UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA

TESIS DOCTORAL

New methologies for studying carbohydrate-protein interacions by nuclear magnetic resonance

Nuevas metodologías para el estudio de la interacción carbohidrato-proteína empleando la resonancia magnética nuclear

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Beatriz Fernández de Toro Ronda

Directores

Francisco Javier Cañada Vicinay
María Ángeles Canales Mayordomo
Jesús Jiménez Barbero

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

NUEVAS METODOLOGÍAS PARA EL ESTUDIO DE LA INTERACCIÓN CARBOHIDRATO-PROTEÍNA EMPLEANDO LA RESONANCIA MAGNÉTICA NUCLEAR.

NEW METHODOLOGIES FOR STUDYING CARBOHYDRATE-PROTEIN INTERACTIONS BY NUCLEAR MAGNETIC RESONANCE.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

BEATRIZ FERNÁNDEZ DE TORO RONDA

DIRECTOR

FRANCISCO JAVIER CAÑADA VICINAY
MARÍA ÁNGELES CANALES MAYORDOMO
JESÚS JIMÉNEZ BARBERO

FACULTAD DE FARMACIA

UNIVERSIDAD
COMPLUTENSE
MADRID

NUEVAS METODOLOGÍAS PARA EL ESTUDIO DE LA INTERACCIÓN CARBOHIDRATO-PROTEÍNA EMPLEANDO LA RESONANCIA MAGNÉTICA NUCLEAR.

NEW METHODOLOGIES FOR STUDYING CARBOHYDRATE-PROTEIN INTERACTIONS BY NUCLEAR MAGNETIC RESONANCE.

TESIS DOCTORAL

BEATRIZ FERNÁNDEZ DE TORO RONDA

DIRECTORES:

FRANCISCO JAVIER CAÑADA VICINAY
MARÍA ÁNGELES CANALES MAYORDOMO
JESÚS JIMÉNEZ BARBERO

MADRID, 2020

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

NUEVAS METODOLOGÍAS PARA EL ESTUDIO DE LA INTERACCIÓN CARBOHIDRATO-PROTEÍNA EMPLEANDO LA RESONANCIA MAGNÉTICA NUCLEAR.

NEW METHODOLOGIES FOR STUDYING CARBOHYDRATE-PROTEIN INTERACTIONS BY NUCLEAR MAGNETIC RESONANCE.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

BEATRIZ FERNÁNDEZ DE TORO RONDA

DIRECTOR

FRANCISCO JAVIER CAÑADA VICINAY
MARÍA ÁNGELES CANALES MAYORDOMO
JESÚS JIMÉNEZ BARBERO

AGRADECIMIENTOS

ACKNOWLEDGEMENTS

En primer lugar, quiero agradecer a mis directores de Tesis el darme la oportunidad de realizar estos años de doctorado. Javier, Jesús y Ángeles, muchísimas gracias por el tiempo dedicado, la paciencia infinita y todo lo que me habéis transmitido.

Quiero recordar a todas las personas con las que he coincidido en el laboratorio a lo largo de estos años, tanto en Madrid como en Bilbao. Gracias por todo lo que me habéis aportado, sin vosotros no hubiese sido lo mismo.

I would also like to thank Jim Paulson for welcoming me in his laboratory during my secondment period.

Por último, agradecer a mi familia y a mis amigos por apoyarme en este camino, que no siempre ha sido fácil. Gracias por los ánimos y por confiar en mí cuando yo no lo he hecho.

CONTENTS

ABBREVIATIONS

A
AA: auxiliary activities
Ala: alanine
Asn: asparagine
Asp: aspartic acid

B
BGL: beta-glucosidase
BME: mercaptoethanol

C
CAZy: carbohydrate-active enzymes
CE: carbohydrate esterase
CMP: cytidine mono-phosphate
CORCEMA: complete relaxation and conformational exchange matrix
CSP: chemical shift perturbation

D
DBrT: 3,5-dibromo-1,2,4-triazole
DHN: 2-(6-hydroxynaphthyl)
DNA: deoxyribonucleic acid
DOSY: diffusion-ordered spectroscopy
DTT: dithiothreitol

E
EDTA: ethylenediamine tetraacetic acid
EGCG: epigallocatechin gallate
ESI: electrospray ionization

F
FBDD: fragment-based drug design
FBS: fetal bovine serum
Fuc: fucose

G
GAG: glycosaminoglycan
Gal: galactose
GalA: galacturonic acid
GalNAc: N-acetyl-galactosamine
gg: gauche-gauche
GH: glycosyl hydrolase
Glc: glucosa
GlcA: glucuronic acid
GlcNAc: N-acetyl-glucosamine
Glu: glutamic acid

Gly: glycine
gt: gauche-trans
GT: glycosyl transferase

H
HA: hemagglutinin
HBT: 1-hydroxybenzotriazole
HEK: human embryonic kidney
HMBC: heteronuclear multiple bond correlation
HMQC: heteronuclear multiple-quantum coherence
HRMS: high resolution mass spectrometry
HSQC: heteronuclear single-quantum coherence

I
IAV: Influenza A virus
IdoA: iduronic acid
Ig: immunoglobulin
IMAC: immobilized meta-affinity chromatography

L
LacNAc: lactosamine
LBT: lanthanide-binding tag
LPS: lipopolysaccharide

M
Man: mannose
MD: molecular dynamic
MSX: L-methionine sulfoximine

N
NA: neuraminidase
Neu5Ac: N-acetyl-neuraminic acid
Neu5Gc: N-glycolyl-neuraminic acid
NMR: nuclear magnetic resonance
NOE: nuclear Overhauser effect
NOESY: NOE spectroscopy

P
PBS: phosphate buffered saline
PCS: pseudo-contact shift
PDB: protein data bank
PEI: polyethylenimine
PL: polysaccharide lyase
pNP: p-nitrophenyl
PRE: paramagnetic relaxation enhancement

R
RBS: receptor binding site
RCA: Ricinus communis agglutinin
RIP: ribosome inactivating protein

S
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser: serine
SLN: sialyl lactosamine
SNFG: symbol nomenclature for glycans
STD: saturation transfer difference

T
tg: trans-gauche
Thr: threonine
TLC: thin-layer chromatography
TOCSY: total correlation spectroscopy
TRIS: tris(hydroxymethyl)aminomethane
trNOESY: transferred NOE spectroscopy
Trp: tryptophan

U
UDP: uridine di-phosphate

W
WHO: world health organization

X
Xyl: xylose

#
5BTT: 5-(5-bromo-2-thienyl)-1H-tetrazole

RESUMEN

NUEVAS METODOLOGÍAS PARA EL ESTUDIO DE LA INTERACCIÓN CARBOHIDRATO-PROTEÍNA EMPLEANDO LA RESONANCIA MAGNÉTICA NUCLEAR.

Los carbohidratos (glicanos, azúcares, sacáridos) juegan un papel clave en la vida y en la enfermedad. Su amplia presencia en diferentes procesos abarca desde las funciones energética y estructural hasta las bases del reconocimiento molecular de eventos fisiológicos y patogénicos. De hecho, se encuentran involucrados en numerosos eventos biológicos de importancia (interacciones célula-célula, célula-matriz o célula-moléculas).

Por lo tanto, el estudio de los eventos de reconocimiento molecular en los que los azúcares se encuentran involucrados es clave para entender estas interacciones, y la elucidación estructural y los estudios conformacionales son cruciales para descifrar sus mecanismos de interacción, así como las implicaciones biológicas de estos eventos.

Sin embargo, la inherente flexibilidad de los carbohidratos impide su cristalización y por tanto su estudio por difracción de rayos X. Por el contrario, la Resonancia Magnética Nuclear (RMN) permite trabajar en disolución, imitando las condiciones fisiológicas , y por tanto, proporciona información sobre la geometría y la dinámica de los azúcares. Por tanto, la espectroscopía de RMN es la técnica preferida para la caracterización estructural de carbohidratos, siendo una técnica muy potente para entender los procesos moleculares en los que los carbohidratos están involucrados.

En este contexto, se han estudiado en esta Tesis dos sistemas biológicos con el objetivo de entender la conformación y la dinámica de diferentes sacáridos y glicomiméticos en solución, y descifrar las características clave de sus procesos de reconocimiento molecular con sus receptores.

Glicosil sulfóxidos diastereoméricos: análisis conformacional y características del reconocimiento frente a β-galactosidasa de *E. coli*.

El mimetismo de los azúcares es una parte esencial del desarrollo de fármacos y/o sondas moleculares basados en carbohidratos. En este contexto, los tiooligosacaridos son ampliamente reconocidos como sondas importantes para el estudio de diversos procesos metabólicos.

En este capítulo se describe el estudio de un conjunto de disacáridos con grupos sulfóxido. La oxidación del átomo de azufre para obtener el sulfóxido da lugar a la obtención de dos posibles estereoisómeros (*S* y *R*) en el átomo de azufre. El análisis conformacional de estos glicomiméticos se ha llevado a cabo utilizando la RMN. Los datos obtenidos por RMN, apoyados por métodos *ab initio* y simulaciones de dinámica molecular, revelan un comportamiento conformacional distinto dependiendo de la estereoquímica de los sulfóxidos. Además, se ha analizado la interacción de estos glicomiméticos con una enzima modelo, la β-galactosidasa de E. coli. A pesar de ser inhibidores moderados, se ha observado la hidrólisis de los glicomiméticos por RMN, aunque la velocidad de hidrólisis es diferente según la estereoquímica del átomo de azufre. Los datos respaldan que la flexibilidad del isómero *S* juega un papel importante en el proceso de selección conformacional que tiene lugar durante su interacción con la enzima. La combinación de datos experimentales de RMN con procedimientos de modelado ha permitido explicar las diferentes velocidades de hidrólisis.

N-glicanos receptores del virus de la gripe A: nuevas metodologías para el análisis conformacional y la caracterización de las interacciones.

La gripe estacional es una enfermedad respiratoria producida por el virus de la gripe y supone una de las mayores amenazas para la salud. El virus de la gripe presenta dos glicoproteínas en su membrana, la Hemaglutinina (HA) y la Neuraminidasa (NA). La combinación de los diferentes subtipos de HA y NA da lugar a las distintas cepas del virus.

La HA es la responsable de iniciar la infección al unirse a N-glicanos sialilados que se encuentran en la superficie de las células del tracto respiratorio. El residuo de ácido siálico está unido a una unidad de N-acetil lactosamina mediante un enlace α2-6 en el caso de los humanos, mientras que en las especies aviares el enlace es α2-3. La mutación de sólo dos aminoácidos permite a los virus aviares cambiar su especificidad del enlace α2-3 al α2-6, permitiendo la infección de humanos y por tanto iniciar una pandemia.

En este contexto, el trabajo descrito en este capítulo se centra en aplicar nuevas metodologías de RMN para obtener información detallada de los epítopos de unión clave requeridos para el reconocimiento por la HA, con el objetivo final de tener un mejor conocimiento de las bases moleculares de la infección del virus de la gripe.

Por lo tanto, se ha realizado el análisis conformacional de N-glicanos relacionados con el virus de la gripe A, así como los estudios de interacción con dos cepas distintas de HA utilizando la RMN.

Aprovechando la aproximación de RMN paramagnética, recientemente desarrollada, se ha obtenido la caracterización completa de la estructura y dinámica de los N-glicanos de estudio. El subsecuente análisis de los datos ha revelado que existen marcadas diferencias en el comportamiento conformacional de los N-glicanos dependiendo de la longitud de las ramas. Además, los estudios de interacción con dos cepas de HA han puesto de manifiesto que, mientras que para la cepa HK/68 no se observan diferencias en la interacción, para la cepa Vic/11 se observa un epítopo mucho más extendido.

Por lo tanto, en esta Tesis se ha realizado un amplio uso de diferentes métodos de RMN para obtener información estructural sobre las interacciones de sistemas proteína-carbohidrato de relevancia biológica y médica. Además, en un contexto relacionado, se ha llevado a cabo un exhaustivo procedimiento de caracterización mediante RMN para elucidar la estructura de una variedad de moléculas que contienen carbohidratos, obtenidas por síntesis enzimática.

ABSTRACT

NEW METHODOLOGIES FOR STUDYING CARBOHYDRATE-PROTEIN INTERACTIONS BY NUCLEAR MAGNETIC RESONANCE.

Carbohydrates (glycan, sugars, saccharides) play a key role in life and disease. Their wide presence in different processes comprises from the structural and energetic function to the basis of the molecular recognition of physiological and pathogenic events. Indeed, they are involved in many biological events of paramount importance (i.e. cell-cell, cell-matrix or cell-molecules interactions).

Therefore, the study of the molecular recognition events in which sugars are involved is a key task for understanding these interactions, and the structural elucidation and the conformational studies of saccharides are crucial to unravel their interaction mechanisms as well as the biological implications of these events.

However, the inherent flexibility of carbohydrates precludes their crystallization, hampering the study by X-ray diffraction. In contrast, Nuclear Magnetic Resonance (NMR) permits to work in solution, mimicking the physiological conditions, and thus provides information on sugar geometries and dynamics. Therefore, NMR spectroscopy is the preferred technique for the structural characterization of carbohydrates, becoming a powerful tool for the understanding of the molecular processes in which carbohydrates are involved.

In this context, two biological systems have been studied in this Thesis with the aim to understand the conformation and dynamics of different saccharides and glycomimetics in solution and to unravel key features of their molecular recognition processes with their receptors.

Diastereomeric glycosyl sulfoxides: conformational analysis and recognition features versus *E. coli* β-galactosidase.

Saccharide mimicry is an essential part of the development of sugar-based drugs and/or molecular probes. In this context, thiooligosaccharides have been largely recognized as important probes for the study of diverse metabolic processes.

The study of a set of sulfoxide disaccharides is described in this chapter. The oxidation of the sulfur atom to obtain a sulfoxide results in the obtention of two possible enantiomers (*S* and *R*). The conformational analysis of these glycomimetics has been deciphered using NMR. The obtained NMR data, supported with *ab initio* methods and molecular dynamic simulations, revealed a different conformational behavior depending on the stereochemistry of the sulfoxides. Moreover, the interaction of these glycomimetics with a model enzyme, *E. coli* β-galactosidase, has been analyzed. Despite being moderate inhibitors, hydrolysis of the glycomimetics has been observed by NMR, although the hydrolysis rates were different, depending on the stereochemistry at the sulfur atom. The data supported that the flexibility of the enantiomer *S* plays an important role in the conformational selection process that takes place in its interaction with the enzyme. The combination of NMR experimental data with modeling procedures has allowed explaining the different rates of hydrolysis.

N-glycans receptors of the influenza A virus: novel methodologies for the conformational analysis and characterization of the interactions.

Seasonal flu is a respiratory disease provoked by influenza virus and is one of the major threats to human health. The influenza virus presents two glycoproteins on the virus membrane, Hemagglutinin (HA) and Neuraminidase (NA). The combination of the different subtypes of HA and NA leads to different virus strains.

HA is responsible for initiating the infection by binding to sialylated glycans located on the surface of our cells of the respiratory tract. The sialic acid residue is linked to a N-acetyl lactosamine unit by an α2-6 linkage in the case of humans, while for avian species the linkage is α2-3. The mutation of only two amino acids permits the avian viruses to switch the specificity from α2-3 linkage to α2-6, permitting the infection of humans and therefore the possibility of initiating a pandemic infection.

In this context, the work described in this chapter has been focused on applying novel NMR methodologies to obtain detailed information on the key glycan binding epitopes required for HA recognition, with the final aim of better understanding the molecular basis of influenza virus infection.

Thus, the conformational analysis of key N-glycans related to influenza A virus as well as their interaction studies with two different HA strains have been performed using NMR.

Taking advantage of the recently developed paramagnetic-based lanthanide approach, the full characterization of the targeted N-glycans has been obtained. The subsequent data analysis has revealed that there are marked differences in the conformational behavior of the N-glycans depending on the length of the branches. Moreover, the interaction studies with two HA strains have permitted to uncover that while for the HK/68 HA strain no differences are observed in the binding of the different N-glycans, the Vic/11 strain presents a much more extended epitope.

Therefore, an extensive use of different NMR methods has been performed in this Thesis to gain structural insights into the interaction of protein-carbohydrate systems of biological and medical relevance. Additionally, in a related context, an exhaustive NMR-based characterization procedure has been carried out to elucidate the structure of a variety of sugar-containing molecules obtained by enzymatic synthesis.

CHAPTER 1

INTRODUCTION

1.1. CARBOHYDRATES

1.1.1. General aspects

Glycoscience is focused on the study of the structure, biosynthesis, biology and evolution of saccharides (also called carbohydrates, sugars and glycans).[1]

The chemistry and biology of carbohydrates started as an important matter of study in the first decades of the past century, being considered as a source of energy and presenting structural functions. In fact, the relevance of the field was manifested by the award of several Nobel laureates in the field of chemistry and physiology or medicine.[2]

However, it has been more recently when it has been recognized that carbohydrates (oligosaccharides, polysaccharides and glycoconjugates) are involved in many biological events (i.e. cell-cell, cell-matrix or cell-molecules interactions) of paramount importance for life and disease. Obviously, this makes the study of their molecular recognition a key task for understanding the response to these interactions. These biomedical and biological processes depend on the chemical and structural diversity of the glycan. For instance, N-glycans have a functional role on cell signaling and adhesion in cancer,[3] sialic acid-containing saccharides are involved in cellular communication on the cell surface and can be exquisitely targeted by siglecs[4] to mediate tumoral, allergy and immune processes.[5,6] Other sugar-binding proteins (lectins), as galectins, specifically recognize N-acetyllactosamine-containing glycans and play a dominant role on regulating pathogen recognition and inflammatory response.[7]

On this basis, it has been in the last three decades when new methodologies have been developed to address these key questions, when the *glycoscience* field has emerged, from glycochemistry to glycotechnology and glycomedicine, passing through glycobiology.[8,9]

1.1.2. Structure of carbohydrates

1.1.2.1. Monosaccharides

The smallest units of carbohydrates are the monosaccharides. They are constituted by poly-hydroxylated carbon chains with a carbonyl group ketone (ketoses) or an aldehyde (aldoses). The huge potential of variety of carbohydrates begins at this level, where different substituents of every carbon convert them in stereogenic centers. This means that, with the subtle change of a hydroxyl group, a different saccharide is obtained.

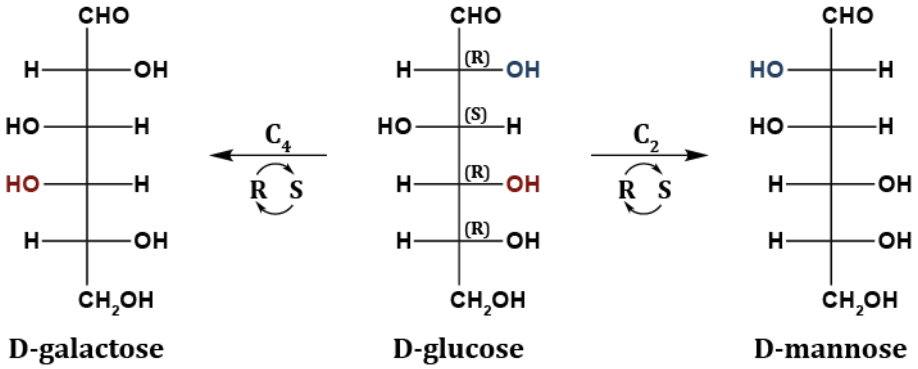


Fig 1. Schematic representation of D-glucose and the epimers D-galactose and D-mannose.

The structural variability increases when the monosaccharides are studied in solution due to the cyclization of the carbohydrates. In the case of hexoses, this cyclization leads to five (furanose) or six (pyranose) carbon membered rings.

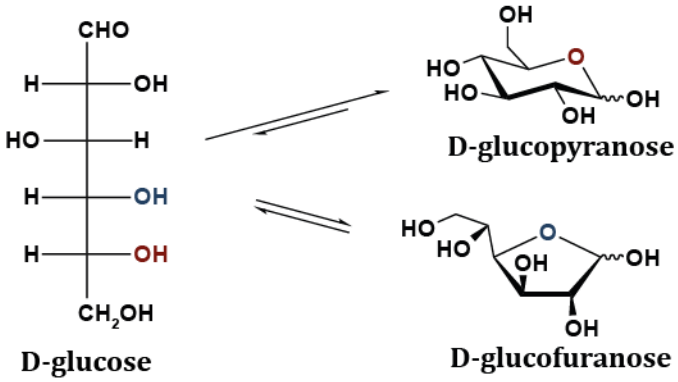


Fig 2. Schematic representation of D-glucose in the open form and pyranose and furanose cycles.

A new asymmetric center is originated when the cyclization occurs, the anomeric position. Considering the orientation of the hydroxyl group at the anomeric carbon, axial or equatorial respect to the plane containing the sugar ring, the α and β

anomers, respectively, are determined. Due to the anomeric effect,[10] the α anomer is favored in the anomeric equilibrium in the case of pyranose rings.

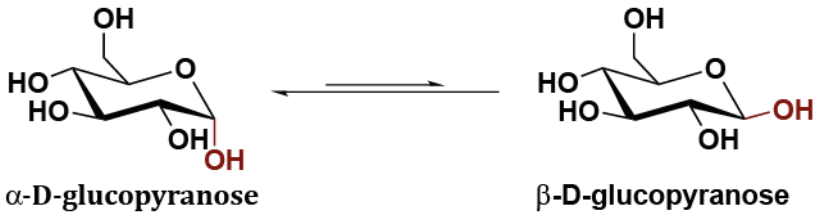


Fig 3. Schematic representation of the α and β anomers of D-glucose.

Moreover, a new source of variability is introduced related to the flexibility of the pyranoses. This is of relevance importance in the molecular recognition of carbohydrates, since it determines the orientation of functional groups in the three-dimensional space.[11]

For the common pyranose rings, the regular geometry is one of the possible chair conformations ($^4C_1$, $^1C_4$). However, different minor geometries can be populated,[12] specially referred to transition states.[13,14] This fact has relevance in the enzymology field, where the distortion of carbohydrates by the enzyme may be a key factor in the reactions.[15-18] Therefore, besides the chair conformation, other minor geometries as envelope, skew or boat conformations are also possible.

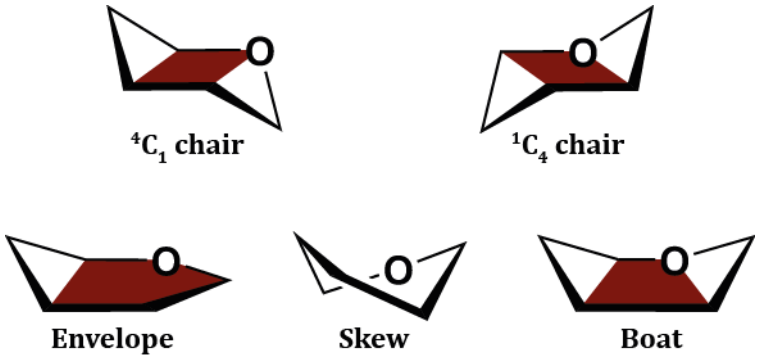


Fig 4. Representation of various geometries adopted by monosaccharide units. The energetically favored chair conformations are labelled as $^4C_1$ and $^1C_4$ depending on the positions of carbons 1 and 4 respect to the ring plane.

Finally, the variability at the simplest complexity level of carbohydrates, the monosaccharides, relies on the possible modifications at the hydroxyl groups, such as acetylation, sulfonation, dihydroxylation and oxidation, among others.

The most common monosaccharides that are found in higher organisms are shown in figure 5.

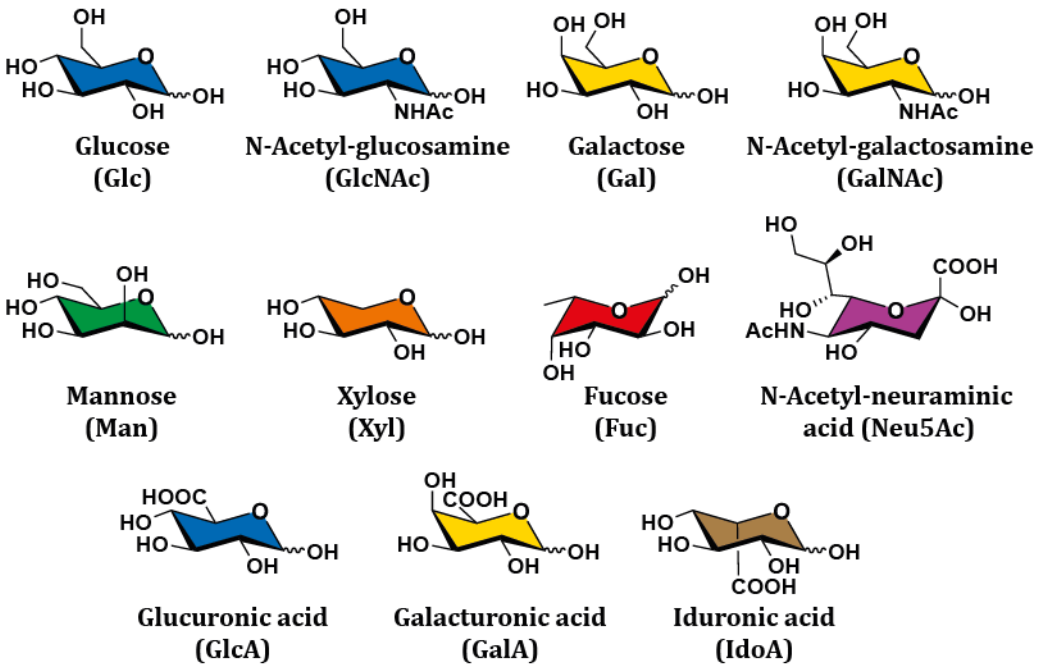


Fig 5. Structure of common monosaccharides found in higher organisms, colored according to the Symbol Nomenclature for Glycans (SNFG).[19] Neu5Ac is the most common form of sialic acid.

1.1.2.2. Disaccharides

The condensation reaction (glycosylation) of two monosaccharides renders a disaccharide, connected by a glycosidic linkage. At this point, the variability is higher even than at the monosaccharide level, given that the monosaccharide at the non-reducing end can adopt both α/β configuration and the multiple available positions of glycosylation.

However, although just around ten monosaccharides are commonly found in higher organisms (figure 5), the wide variability introduced by the formation of glycosidic linkages can be shown with the multiple disaccharides that can be formed with just two glucose units (figure 6).[20,21]

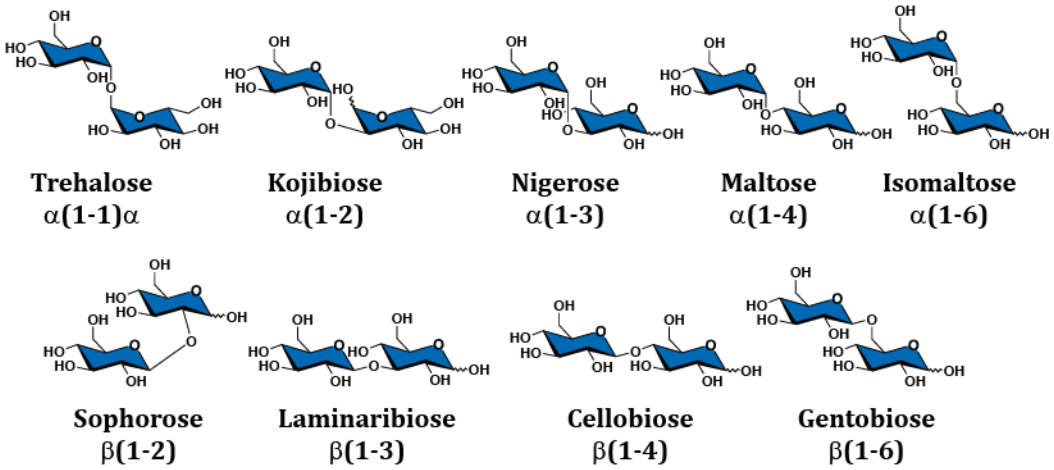


Fig 6. Disaccharides obtained by the linkage at different positions of two glucose units.

The glycosidic linkage is the most flexible part of a disaccharide unit and defines its three-dimensional shape. The conformations around this linkage are defined by the torsion angles Φ (H_1-C_1-O-C_x), Ψ (C_1-O-C_x-H_x) and for 1-6 linkages, ω (O_6-C_6-C_5-C_4).

Torsion angles Φ and Ψ relate the orientation of the monosaccharides to each other (figure 7). The *syn*-Φ/*syn*-Ψ, *syn*-Φ/*anti*-Ψ, and *anti*-Φ/*syn*-Ψ conformational states are given by dihedral values of ±60° for the *syn* conformer and 180° for the *anti*-conformer.[22]

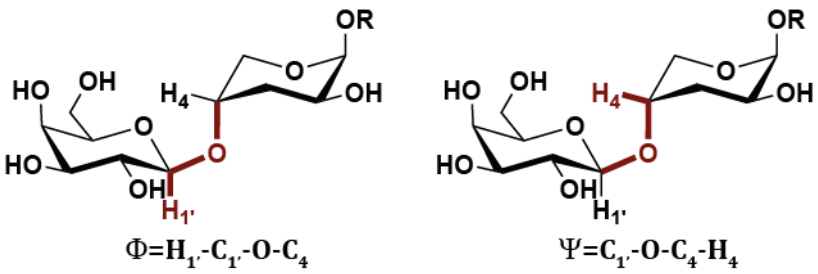


Fig 7. The torsion angles around the glycosidic linkage that define the different conformers. Analogous disaccharide to the sulfoxides studied in chapter 3.

For 1-6 linkages, the ω torsion describes the possible gg (*gauche-gauche*), gt (*gauche-trans*) and tg (*trans-gauche*) rotamers. This terminology refers first to the O_6-O_5 torsion angle followed by that around O_6-C_4.

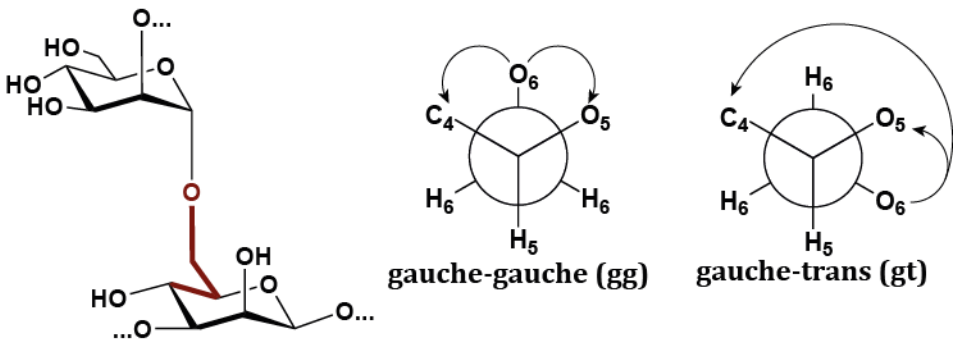


Fig 8. The rotamers defined by ω torsion angle found for the N-glycans core of the complex type N-glycans studied in chapter 4. In this case, the ω torsion angle is defined for the Manα (1-6) Manβ linkage. The Newman projections for both gg and gt rotamers are shown.

Then, a new degree of complexity and variety is open for carbohydrates, the 3D conformational variety.

1.1.2.3. Glycomimetics

Sugar mimicry constitutes another source of complexity in the carbohydrate field, being essential for the development of carbohydrate-based therapeutics.[23] Glycomimetics are sugar scaffolds bearing modifications that change the saccharide properties while still resembling the natural carbohydrate.[24,25]

The glycosidic linkage can be modified replacing the exocyclic-anomeric oxygen atom by another atom type as carbon, nitrogen or sulfur, leading to C-, N- or S glycosyl compounds (figure 9A).[26-29] Moreover, the endocyclic oxygen can also be modified, rendering carba-, amino- or thio- sugars (figure 9B).[27,29-35] Furthermore, deoxy- glycan analogues have been employed to monitor interaction events, assessing the contribution of the specific sugar hydroxyl groups to the binding event (figure 9C).[36,37]

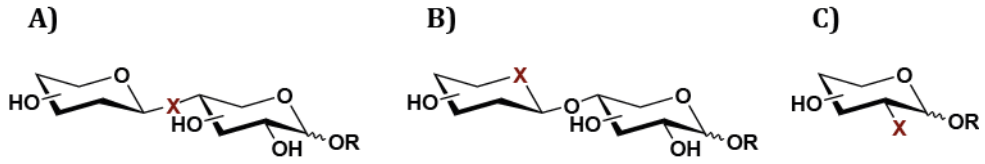


Fig 9. Sugar mimics. General structure of the replacement of **A)** the exocyclic oxygen atom, **B)** the endocyclic oxygen atom, and **C)** deoxy- glycan analogues.

Glycomimetic molecules can disrupt the formation of sugar-protein complexes, interfering in the molecular recognition event and modifying the enzymes activities. In this context, their use as probes of biological processes and for medicinal applications make carbohydrates a highly attractive target for drug design.[25,38,39]

1.1.2.4. Further complexity

Different structures can be obtained by linking monosaccharides through glycosidic linkages, rendering oligosaccharides (less than 12 saccharide units) or polysaccharides (more than 12). The polysaccharides comprise a vast variety of functions as energy storage (glycogen, starch) and the structural composition of plant (cellulose, xylan) and fungal (chitin) cells. Moreover, they are presented at the cell-surface of animals (GAGs) and bacteria (LPS).

Depending on the nature of the carbohydrates and the variety of the glycosidic linkages, not only linear products but also branched structures can be found in nature. Branching is a characteristic of many glycans located on the mammalian cell surfaces. For N-glycans, an asparagine residue is attached covalently to the glycan, followed by any amino-acid (except proline) and serine or threonine: the Asn-X-Ser/Thr triad.[40] All eukaryotic N-glycans share a common core sequence and, depending on the extensions of the branches, three types of N-glycans are obtained: high mannose, complex and hybrid types.

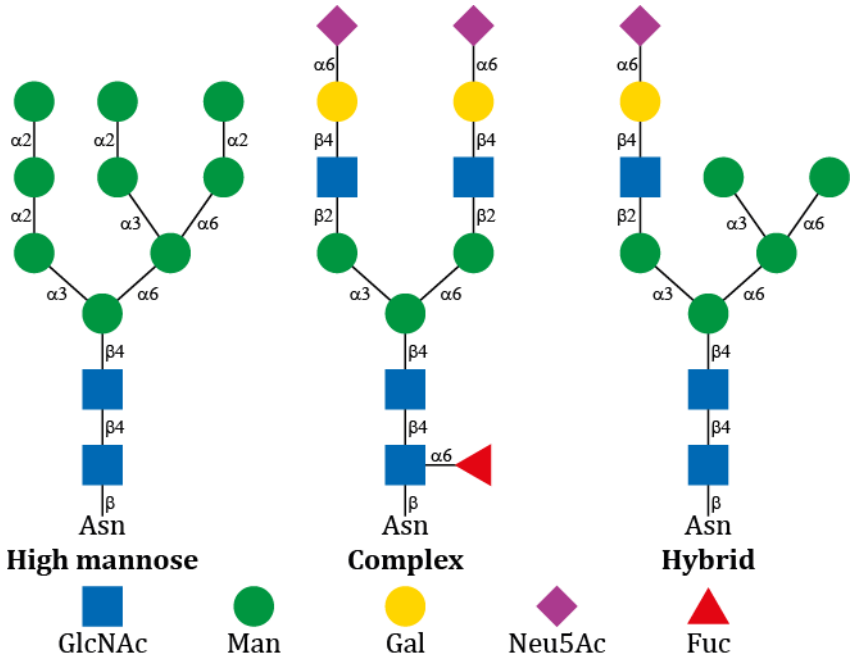


Fig 10. Top: Types of N-glycans. The common core sequence (Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn-X-Ser/Thr) is extended with: Man residues (high mannose type), an initial GlcNAc extension with Galβ1-4GlcNAc (LacNAc) repeats (complex type), Man residues at the Manα1-6 arm and GlcNAc at the Manα1-3 arm (hybrid type). **Bottom:** Saccharide units caption colored according to the Symbol Nomenclature for Glycans (SNFG).[19]

1.2. MOLECULAR RECOGNITION OF CARBOHYDRATES

1.2.1. Biological significance

Carbohydrates constitute one of the most abundant and complex type of biomolecules in nature. All cell surfaces are coated by carbohydrates forming the glycocalyx (figure 11), which represents a physical barrier. Furthermore, glycans are widely expressed as glycolipids and glycoproteins, and glycosylation is one of the most prevalent post-translational modification of proteins.

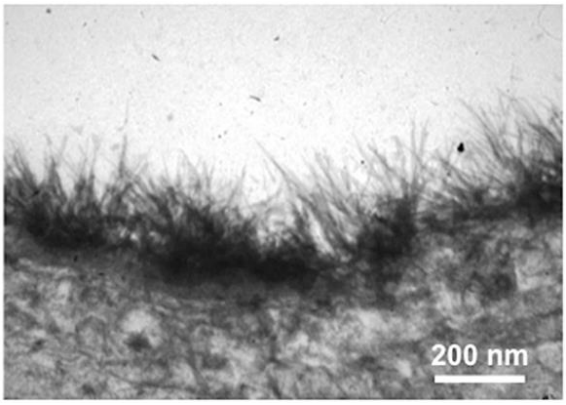


Fig 11. Carbohydrates forming the glycocalyx. (Reprinted from British Journal of Clinical Pharmacology, vol 80, Becker, B. F.; Jacob, M.; Leipert, S.; Salmon, A. H. J. and Chappell, D., 389-402, Degradation of the endothelial glycocalyx in clinical settings: searching for the sheddases, Copyright (2015), with permission from John Wiley and sons).

The well-known and previously mentioned energetic (glycogen and starch in animal cells and plants, respectively) and structural functions of carbohydrates (cellulose and chitin provide strength and rigidity to plants, fungal cell walls and arthropod exo-skeletons) are encompassed in the structural and modulatory roles of glycans.

In addition to these functions, carbohydrates play very important roles in molecular recognition processes by establishing specific interactions with different entities such as enzymes, viruses, bacteria and antibodies among others (extrinsic recognition), as well as their involvement in cell-cell interactions (intrinsic recognition).[41,42]

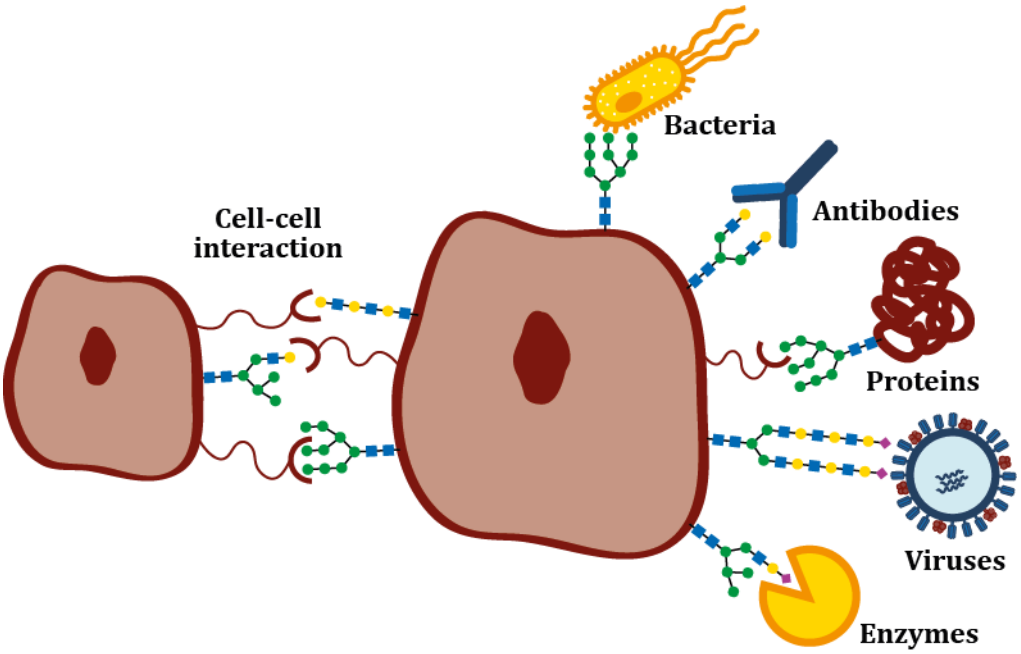


Fig 12. The importance of protein-carbohydrate interactions. Cell-surface carbohydrates are recognized by a diversity of entities and play a key role in cell-cell interactions.

Some examples of the importance of carbohydrates in molecular recognition events as well as other related curiosities are given below:

The ABO blood group:

The discovery of the human blood groups by Karl Landsteiner, who was laureated with the Nobel prize in 1930 "for his discovery of human blood groups",[2] marked a breakthrough in blood transfusions. Until then, the assumption of a universal blood had caused tragic consequences. The subtle difference of presenting an additional Gal or GalNAc unit attached to the ABO antigen precursor (H-antigen) on the red blood cells surfaces encode the different blood group types (figure 13).[43] A recent research in this field has directed its efforts in the enzymatic conversion of the A and B types into the universal O type.[44]

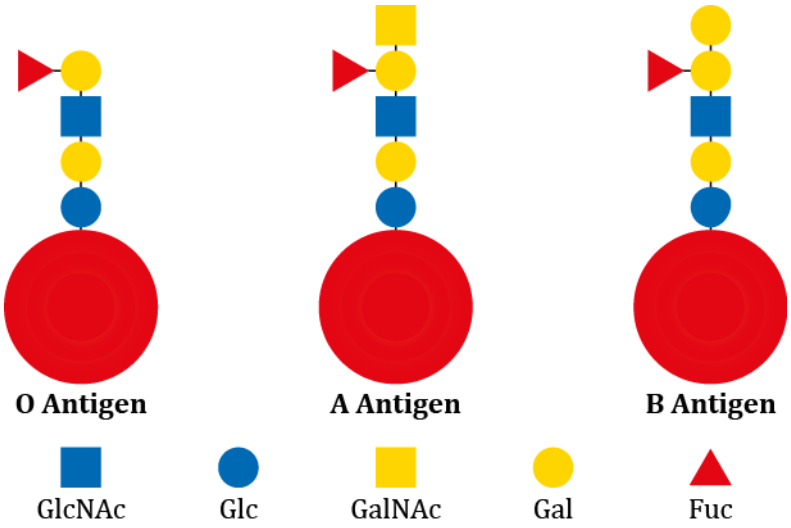


Fig 13. Subtle differences that define the ABO groups.

Fertilization and reproduction:

The glycan recognition is also involved in the sperm-egg interaction and in many steps in the reproductive process such as migration to the site of fertilization, implantation and placental functions.[45-47]

Antigenic epitopes:

The variations in glycosylation can result in strongly antigenic epitopes. For example, the "Lone Star tick" bite (*Amblyomma americanum*) induces the development of IgE antibodies against α-Gal epitopes (not presented in humans). Later exposure to mammalian foods rich in α-Gal motifs, as red meat, make these people suffer reaction in an apparent "red meat allergy".[48,49]

Besides the α-Gal epitope, Neu5Gc is not present in humans. Both represent the major xeno-antigens that need to be overcome in xeno-transplantations into humans. In this context, α-Gal- and Neu5Gc- double null pigs have been generated.[50,51]

Ricin toxin:

Ricin, a R-type lectin (RCA-II), is purified from *Ricinus communis* seeds. Ricin binds to cells by interacting with β-linked Gal/GalNAc glycans on the cell surface and is imported into the endosomes. Its toxicity derives from its action as a type II ribosome inactivating protein (RIP-II).[52,53]

Ricin is defined as a biological category B agent and it has been used as biological weapon. One of the most famous cases was the murder of the Bulgarian dissident Georgi Markov by the KGB. He was assassinated in 1978 in London after the injection of a pin-sized metal pellet using a poisoned umbrella.[54] Our group has extensively studied the recognition features of this toxic lectin.[55-58]

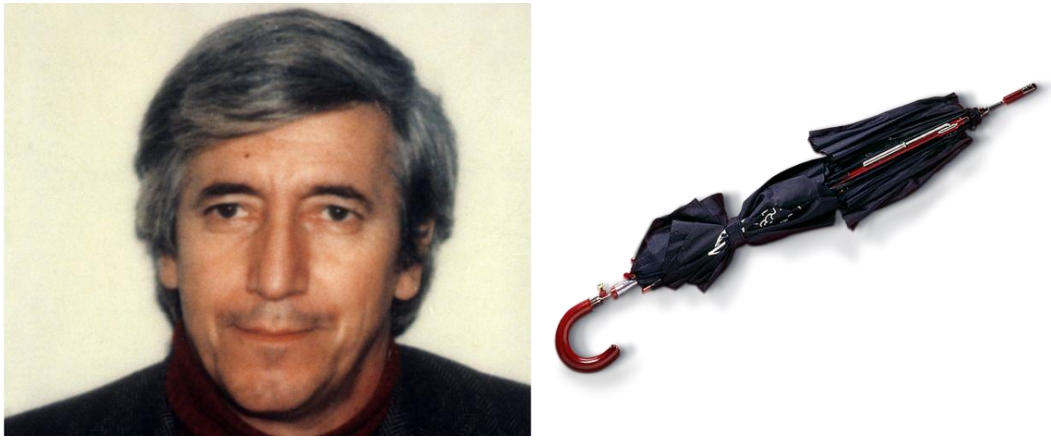


Fig 14. Left: Portrait of Georgi Markov (taken from BBC). **Right:** The umbrella used in the murder, exposed in the spy museum of Washington. Courtesy of the International Spy Museum.

In this context, the study of the 3D structure of carbohydrates and the characterization of their recognition features are essential for a thorough understanding of many vital processes and their modulation.[59,60]

1.2.2. Carbohydrate-protein interaction

Proteins are the most common receptors in the biological molecular recognition processes. Carbohydrates and other ligands are recognized by biological partners belonging to different protein families such as enzymes, antibodies, and lectins.

The enzymes that catalyze the synthesis, disruption or modification of sugars receive the name of carbohydrate-active enzymes (CAZymes), most of them included and classified in the CAZy database.[61] Currently, CAZy contains information on five large enzyme classes: glycosyl hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and enzymes displaying auxiliary activities (AAs).[62-65]

Anti-glycan antibodies are generated by the immune system and play a key role in the defense against pathogens, based on their ability to distinguish self-antigens from pathogen-derived antigens.[42]

Finally, lectins are a large family of proteins from different sources that specifically recognize carbohydrates without producing any transformation on them. Lectins (Latin for "select") recognize and bind selectively and specifically to carbohydrate epitopes of glycoproteins, glycolipids and free oligosaccharides. The major biological functions of lectins comprise the trafficking, targeting and clearance of proteins, cell adhesion, and immunity and infection.[66]

1.2.3. Types of interactions

Carbohydrate-protein binding is possible due to the existence of weak intermolecular forces between both species, including polar and non-polar interactions.

1.2.3.1. Hydrogen bonding

The formation of a hydrogen bond involves a hydrogen donor, characterized by the presence of a hydrogen covalently bonded to an electronegative atom and a hydrogen acceptor, also an electronegative atom. This atom provides a lone pair of electrons to be shared with the hydrogen of the donor. The occurrence of hydrogen bonds between sugars and their receptors constitute the most evident interaction in the carbohydrate field, due to the high presence of hydroxyl groups in all saccharides. They play an important role in the molecular recognition events and contribute not only to the affinity but also to selectivity.

These -OH groups can participate as donors and as acceptors, and in some cases, the same group simultaneously display both roles. This phenomenon is known as cooperative hydrogen bonding.

Therefore, carbohydrates establish hydrogen bonds with the side chains of polar residues of proteins through their amine and carbonyl groups.[67]

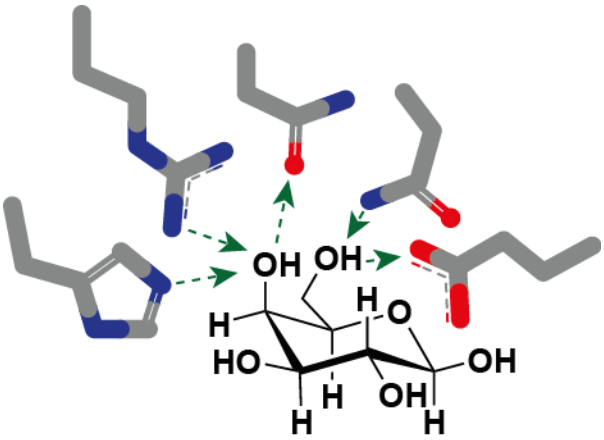


Fig 15. Representation of some possible hydrogen bonding between the -OH groups at positions 4 and 6 of a galactose unit and diverse amino acids.

1.2.3.2. Non-polar interactions

Besides hydrogen bonding, non-polar interactions also contribute to the molecular recognition between biochemical species. In some carbohydrates, the clustering of three or more C-H groups caused by the disposition of hydroxyl groups forms hydrophobic patches on the sugar surface, allowing the establishment of non-polar interactions with hydrophobic epitopes of proteins. The aromatic protein residues are of special interest, creating CH-π interactions with the C-H groups of the saccharide perpendicularly oriented to the ring.[68-70]

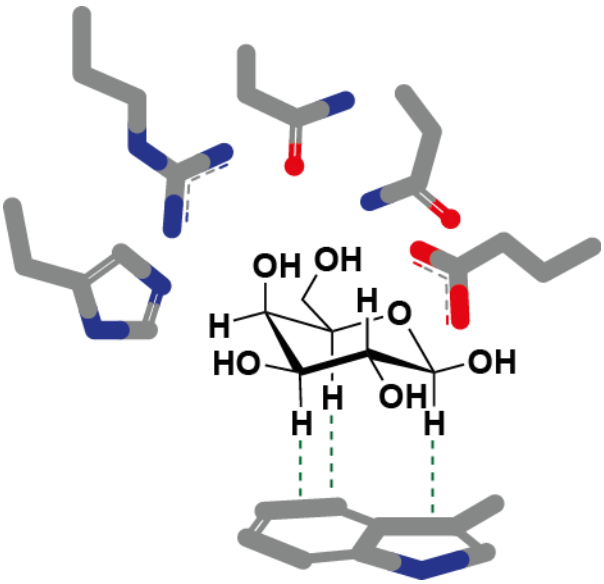


Fig 16. Representation of the possible CH-π interaction between the -CH groups of a galactose unit and a simple aromatic amino acid side chain (Trp).

1.2.3.3. Other interactions

Due to the complexity of the biochemical systems involved in the molecular recognition events, other interactions in addition to polar and non-polar contacts may take place. The electrostatic interaction between charge saccharides, such as sulfated GAGs, and protein residues of opposite charge can be considered in some systems, as well as the involvement of a third partner, such as ions or water molecules. For instance, the coordination of a divalent cation (Ca^{2+}) is required for the binding of carbohydrate ligands by proteins belonging to the C-type family of lectins.[71,72]

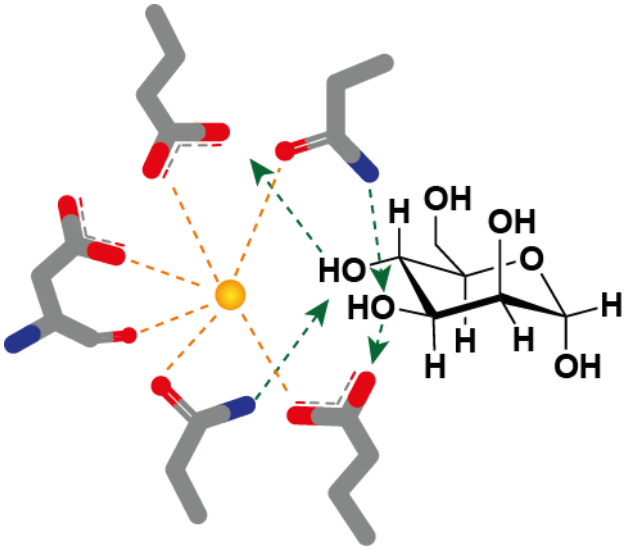


Fig 17. Representation of the involvement of a Ca^{2+} ion in the molecular recognition of mannose by a mannose-binding C-type lectin.

1.3. NUCLEAR MAGNETIC RESONANCE

1.3.1. NMR principles

Nuclear Magnetic Resonance (NMR) is a spectroscopic technique applied in different scientific and technology fields, including chemistry, biology, and medicine. NMR is one of the most important techniques to determine the structure, conformation, dynamic and interactions of molecules.

Since the first observation of proton magnetic resonance (^{1}H NMR) in solids and liquids in 1945,[73,74] NMR has undergone through an incomparable growth, demonstrating to be a disruptive methodology in science. It is noteworthy to mention that six laureates have four Nobel prizes related to NMR: Felix Bloch and Edward Mills Purcell (Nobel Prize in Physics 1952 "for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"), Richard R. Ernst (Nobel Prize in Chemistry 1991 "for his contributions to the development of the methodology of high resolution nuclear magnetic resonance spectroscopy"), Kurt Wüthrich (Nobel Prize in Chemistry 2002 "for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution") and Paul C. Lauterbur and Sir Peter Mansfield (Nobel Prize in Physiology or Medicine 2003 "for their discoveries concerning magnetic resonance imaging").[2]

NMR is a physics phenomenon based on the quantum-mechanic properties of the atomic nuclei.[75,76] The nuclei of the atoms are characterized by a nuclear spin quantum number (I) and only those nuclei presenting $I \neq 0$ are observable in NMR. Such nuclei possess an associated magnetic moment (μ), dependent of the angular momentum (P) and of the magnetogyric ratio (γ), which is a constant characteristic of each nucleus.

$$\mu = \gamma P$$

Equation 1. Relationship between the angular momentum and the magnetic moment of a nuclear spin.

The angular momentum and the magnetic moment are vector quantities, i.e. they have both magnitude and direction. For a given nucleus with spin I, there exist $2I+1$

possible spin states. Therefore, for nuclei with spin I=1/2 such are ^{1}H, ^{13}C and ^{15}N, there are two possible spin states, +1/2 and -1/2, dubbed α and β. The α and β states are degenerated, they have the same energy. However, in the presence of a magnetic field (B_0) the degeneracy is broken due to the interaction of the magnetic moment with the magnetic field. As a result, there are two different energy levels, with a slightly larger fraction of the nuclei presented in the lower energy state, following the Boltzmann distribution.

$$\Delta E = \gamma \hbar B_0$$

Equation 2. Energy difference between spin levels, which is proportional to the strength of the field.

$$\frac{N_{low}}{N_{upp}} = e^{\frac{-\Delta E}{kT}}$$

Equation 3. Boltzmann distribution of the populations.

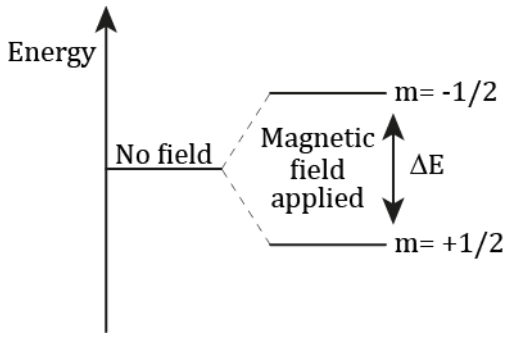


Fig 18. Energy separation between nuclear spin states.

The asymmetric distribution of the nuclear spin in the two states in the presence of a magnetic field is the basis of the NMR experiments. The NMR phenomenon occurs when, by applying an energy (electromagnetic radiation), the spins go from the lower to the higher energy state. Then, the system tends to recover the initial state and the energy released in the transition to the equilibrium is translated into the NMR signals found in the spectra.

The magnetic moment (μ) of each nucleus precesses around the external magnetic field (B_0). The frequency of this precession is the Larmor frequency (ω), which is characteristic of each nucleus because it depends on the magnetogyric ratio (γ) and is equivalent to the energy difference between the two levels.

$$\omega = \gamma B_0 = \frac{\Delta E}{\hbar}$$

Equation 4. Larmor frequency.

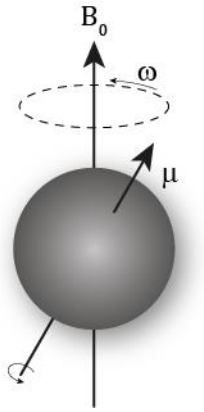


Fig 19. Precession of a nucleus with an associated magnetic moment (μ) around the magnetic field (B_0).

The Larmor frequency for each nucleus depends therefore on the magnetogyric ratio and the strength of the magnetic field. Nevertheless, there exist local perturbations of the magnetic field which are provoked by the surrounding nuclei. As a result, nuclei within different chemical environments precess at slightly different frequencies, even being the same type of nucleus, which permits the use of NMR as a spectroscopy technique. For practical reasons, the use of chemical shifts (δ, expressed in ppm) instead of Larmor frequencies is preferred. This nomenclature permits the comparison of the measurements in spectrometers operating at different external magnetic fields.

$$\delta = \left(\frac{\omega - \omega_{ref}}{\omega_{ref}}\right) 10^6$$

Equation 5. Chemical shift in ppm for the Larmor frequency of a given nucleus and a reference compound (usually TMS).

1.3.2. NMR of carbohydrates

Over the last years, the obtention of fine structural details of carbohydrates and the recognition by receptors have been the goal of many investigations. The advances in structural techniques, such as NMR and X-ray crystallography, have permitted to study different protein-sugar complexes of biological interest at atomic resolution.[77-83]

NMR spectroscopy is the preferred technique for structural characterization of carbohydrates, becoming a powerful tool for a full understanding of the molecular processes in which carbohydrates are involved. The inherent flexibility of carbohydrates precludes the crystallization, hampering the study by X-ray diffraction. In contrast, NMR permits to work in solution, mimicking the physiological conditions, and thus providing information of the geometries and dynamics.[71,79,81]

The NMR experiments provide key information to determine the structure of carbohydrates. Among several experimental restrains, vicinal ^{1}H-^{1}H coupling constants ($^{3}J_{HH}$) and the nuclear Overhauser effect (NOE) are particularly useful.

1.3.2.1. Coupling constants

Martin Karplus (Nobel Prize in Chemistry 2013) described a relationship between the torsion angle value relating vicinal hydrogens (H-X-Y-H) and the $^{3}J_{HH}$ coupling constant.[84]

$$^{3}J_{HH} = A + B\cos\varphi + C\cos^{2}\varphi$$

Equation 6. The standard Karplus equation.

Therefore, the average $^{3}J_{HH}$ for an ensemble of conformations can be estimated from Karplus-type equations and compared to the experimental values. Thus, the characterization of the population distributions in conformational equilibria is possible.[32,34,85,86]

1.3.2.2. The Nuclear Overhauser Effect (NOE)

The study of the 3D molecular geometry of carbohydrates becomes more complex as the number of saccharide units increases. The access to information regarding the spatial relationship between nuclei is a powerful tool to elucidate the conformational disposition of carbohydrates.

The NOE is defined as the change observed in the NMR signal intensity of a nucleus when another nucleus is perturbed by a selective excitation. These two nuclei should be close enough to share a dipolar coupling (not the above-mentioned scalar coupling, which is through bond). Therefore, the NOE depends on the distance and on the molecular motion of the interproton vector.[87]

$$I_{NOE} \approx \langle 1/r^6 \rangle f(\tau_c)$$

Equation 7. Dependence of the NOE intensity on the distance (r) and correlation time (τ_c), which describes the motion of the interproton vector.

The distance dependence of the NOE (r^{-6}) implies that only nuclei close in space (within 5-6 Å) develop NOE. Therefore, NOESY spectra contain information about interproton distances, which can be translated into conformational information.

1.3.2.3. Molecular recognition: The bound state

In addition to the NMR methods suitable for studying carbohydrates in solution, the recognition event of protein-ligand complexes can also be studied by a variety of NMR experiments. These methods permit to study the key features of the binding process, tackling the analysis from the ligand- or receptor-perspective, using specific tailor-made NMR experiments.

Binding events are defined, according to thermodynamics, as an equilibrium between the free species and the bound complex.[88]

$$R + L \underset{k_{off}}{\rightleftharpoons} RL$$

Equation 8. Equilibrium at the binding event.

k_{on} is the association rate and k_{off} the dissociation rate, and the dissociation constant (K_D) can be described as follow:

$$K_D = \frac{[R][L]}{[RL]} = \frac{k_{off}}{k_{on}}$$

Equation 9. Expression of the dissociation constant.

Ligands (small molecules, sugar in this case) and receptors (macromolecules) have different motional properties. Small molecules present fast Brownian motion, slow relaxation, fast diffusion and positive NOEs. In contrast, these properties display opposite values for the macromolecules. When a small molecule binds to a macromolecule, upon a recognition event, it acquires the motional properties of the macromolecule and therefore, the NMR properties of the ligands are altered.[89]

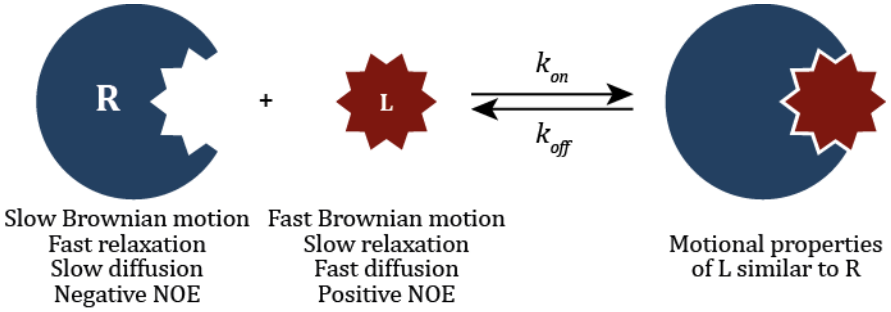


Fig 20. NMR properties of receptors and ligands.

Therefore, the study of the molecular recognition event can be addressed from two different perspectives, either observing the ligand or the receptor NMR properties. Fittingly, both approaches provide complementary information.[83]

1.3.2.3.1. Ligand-based methods

This approach is based on the monitorization of the NMR parameters of the ligands in the absence and presence of the macromolecules. Saturation Transfer difference (STD) and transferred NOE spectroscopy (trNOESY) are used to detect the changes in the motional properties of the ligands upon binding. The size of the protein is not an issue, in contrast to the receptor-based methods, and neither isotopic labelling nor high quantities of protein are required. In fact, the larger the macromolecule is, the more sensitive the ligand-based experiments are. However, some drawbacks are

presented when using these methods. Complexes with relatively fast dissociation kinetics are required, which are usually related to weak or moderate affinities (K_D in the 10^{-8} to 10^{-2} M range). Since these methods rely on the exchange-mediated transfer of bound state information to the free state, if the residence time of the ligand in the receptor binding site is too high, fast relaxation will operate, killing the generated information before the transfer process takes place. Thus, no information on the binding event will be obtained.

1.3.2.3.1.1. Saturation Transfer Difference (STD)

The STD-NMR experiment is a powerful tool for identifying ligand-receptor interactions. It is based on the transfer of magnetization from the receptor, which is selectively irradiated by a train of radiofrequency pulses, to any ligand bound to it.

The experiment involves the comparison of two ^{1}H-NMR spectra of the same sample, acquired either with (on-resonance spectrum) or without (off-resonance spectrum) saturation of the protein.[88,90] In the on-resonance experiment, a saturation pulse is applied to the protein signals, usually at the aliphatic region (between δ -2 and 0.5 ppm) or the aromatic region (between δ 6 and 8 ppm) if there are not aromatic protons in the ligand structure.[77] The saturation then propagates from the irradiated protons to the rest of the protein (^{1}H-^{1}H cross relaxation mechanisms) and from there to any ligand in direct contact to the protein (< 4-5 Å). This renders a ^{1}H-NMR spectrum where the signal intensities of the saturated protons of the ligand decrease depending on its distance to the receptor. On the other hand, the off-resonance experiment is obtained by irradiating the sample at spectral regions where signals of the macromolecule and the ligands do not appear (e.g. δ 100 ppm). The difference between the off- and the on-resonance spectra yields the STD spectrum, where only the signals of the ligand protons that are involved in the interaction with the macromolecule do appear.

The analysis of the STD NMR signals permits to screen a set of compounds, and it is widely employed in the pharmaceutic industry for screening and for fragment-based drug design (FBDD).[91] Moreover, the STD NMR experiment provides

information on the ligand binding epitope, since the protons that are closer to the protein become more saturated than those that do not participate directly in the interaction.[92]

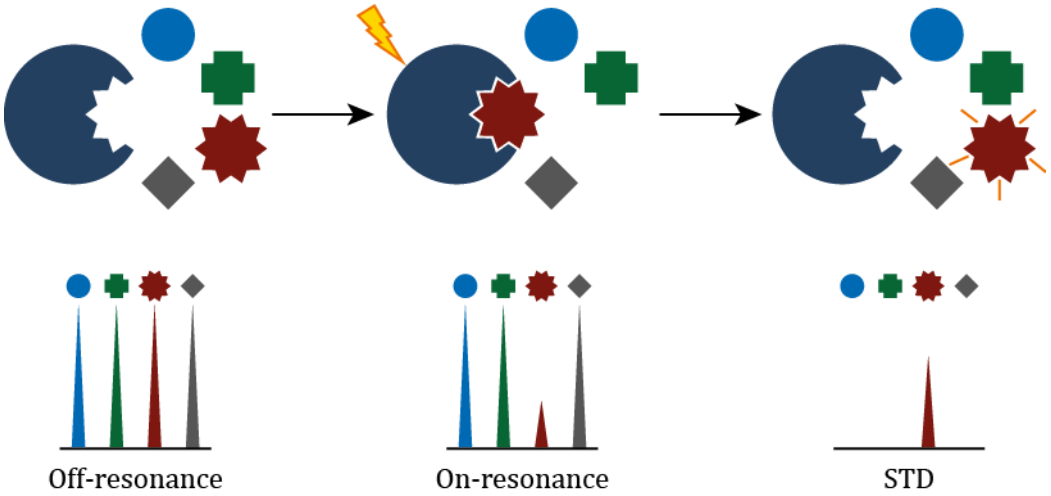


Fig 21. Top: Selective transfer of magnetization from the irradiated protein to the ligand that binds. **Bottom:** Simulated off-resonance, on-resonance and STD NMR spectra, in which saturation of the binding molecule leads to a decrease of its NMR signal in the on-resonance spectrum. In the STD NMR spectrum, only the signals of the ligand that binds appears.

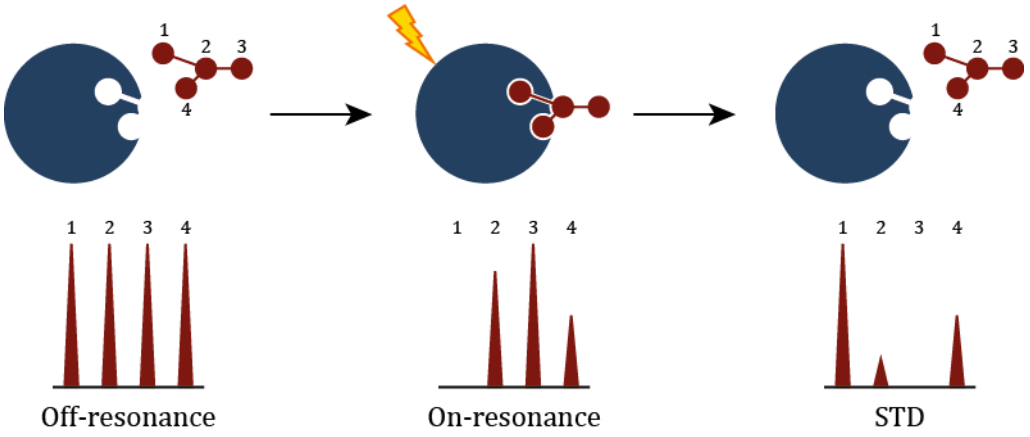


Fig 22. Top: Selective transfer of magnetization from the irradiated protein to the different parts of the ligand. **Bottom:** Simulated off-resonance, on-resonance and STD NMR spectra. In this case, the epitope binding of the ligand can be obtained from the different STD NMR intensities. Closer nuclei receive higher saturation from the protein.

1.3.2.3.1.2. Transferred NOESY (trNOESY)

The trNOE spectroscopy has been widely used to study receptor-ligand interactions since its introduction by Bothner-By in the early 1970s.[93,94] The trNOESY experiment is a regular 2D NOESY experiment in which the cross-peaks of the ligand

change their signs, due to the dramatic variation in the ligand motion properties upon binding to the receptor. As previously mentioned, small molecules present positive NOEs, while macromolecules have negative NOEs. As a result, non-binders show no NOE sign change in the presence of a putative receptor and their NOEs remain positive, while binders now show negative NOE cross-peaks (trNOEs). The observed NOEs are a contribution of both the free and the bound ligand (the experiments are acquired in a large excess of ligand), but the cross-relaxation rate of bound ligand is much larger than that in the free state. Therefore, the observed trNOEs show a predominance of the contribution arising from the bound form.[77] As a result, the trNOESY experiment not only permits to detect interactions, but also provides information on the conformation of the ligand in the bound state.[95,96]

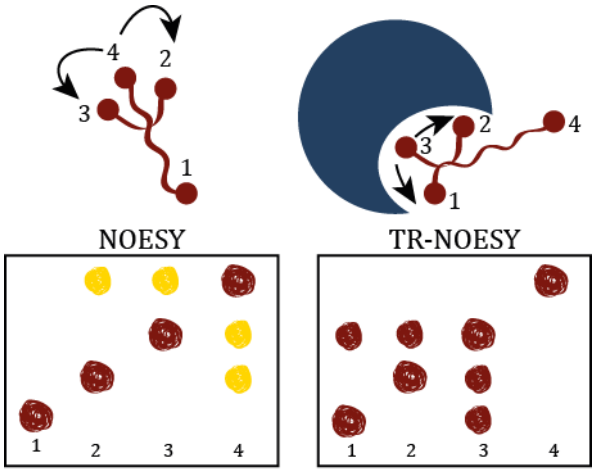


Fig 23. Left: Schematic representation of the NOESY spectrum of a free ligand (positive sign). Cross-peaks and diagonal peaks have different sign. **Right:** Schematic representation of the trNOESY spectrum of a ligand complexed with a receptor. Cross-peaks and diagonal peaks have the same sign (negative sign), indicating binding. The information provided in the spectra permits to detect conformational variation.

1.3.2.3.2. Receptor-based methods

These methods are based on the monitorization and comparison of the receptor signals in the absence and presence of the ligands. In this case, the binding site can be studied or even discovered by identifying specific protein NMR signals. However, some disadvantages accompany these methods. The target proteins should be relatively small (<35 kDa) and high quantities of soluble and non-aggregated proteins must be expressed and purified. Moreover, the expression systems must

permit enrichment with stable NMR active isotopes, needed for the assignment of the protein NMR signals, which requires stable samples for the time required by the NMR experiments.

1.3.2.3.2.1. Chemical Shift Perturbations (CSP)

One of the most frequently used NMR methods relies on the analysis of the perturbation of the chemical shifts (CSP) of the protein upon addition of the ligand. Given the high overlapping that the spectra of a protein present in the proton dimension, most of the experiments used for studying the binding event are 2D experiments, specially the ^{1}H-^{15}N and ^{1}H-^{13}C HSQC techniques.

The monitorization of the CSP of a protein in a HSQC spectrum, as a result of a change in the chemical environment of the amino-acids involved in the molecular recognition, provides information on the residues involved in the interaction, the receptor binding epitope. The ^{13}C labelling is more expensive and the ^{1}H-^{13}C HSQC spectra are more complex to analyze; therefore, the ^{1}H-^{15}N HSQC is the most commonly used method.[88,97]

Moreover, for simple systems, CSP can also be studied in ^{1}H-^{1}H 2D experiments, such as NOESY, using a non-labelled receptor. In this case, information on the conformation of the protein both in the free and bound states can be obtained, elucidating if there are conformational changes when the molecular recognition process takes place.[79]

1.3.2.3.2.2. Other NMR experiments

Other experiments may also be useful for studying the molecular recognition event from the receptor's point of view. For instance, selective ^{13}C-labelling of amino-acids can provide key information through ^{1}H-^{13}C experiments.[98] The use of ^{19}F-labelled proteins may permit obtaining clean 1D ^{19}F-NMR spectra that allow monitoring perturbations on NMR parameters upon binding.[99]

Moreover, nitroxide spin-labelled carbohydrates[100] and carbohydrates bearing lanthanide binding tags[101] have been employed for mapping carbohydrate binding sites by measuring the loss of NMR signals intensities due to the paramagnetic relaxation enhancement (PRE) or their variations due to the presence of pseudo contact shifts (PCS). These methodologies have opened new avenues in the field that will be described below.

1.3.3. The frontier: paramagnetic NMR

The standard NMR parameters that provide structural information, previously described, are chemical shifts (δ), scalar couplings (*J*), and Nuclear Overhauser Effects (NOEs). However, the narrow chemical shift window presented by glycans in addition to the restrained local structural information obtained by the previously mentioned NMR studies, forested the necessity of progressing beyond the standard NMR techniques and develop novel disrupting methodologies. In this context, the use of paramagnetic nuclei, including lanthanides attached to the glycans through chelating units has permitted to go a step further.

The presence of a paramagnetic lanthanide induces two main effects on the surrounding nuclei: Pseudo-Contact Shift (PCS) and Paramagnetic Relaxation Enhancement (PRE).[102] The relaxation induced by paramagnetic nuclei is very strong. The PRE decreases with r^{-6}, where r is the distance between the metal ion and the nuclear spin. The PRE induces a dramatic line broadening in the signals of the NMR active nuclei that are close to the paramagnetic center and may frequently be hidden under the noise level. These features have been employed to assess interactions in the glycan-receptor interaction field.[103]

On the other hand, PCS are the result of a dipolar interaction between the unpaired electron of the lanthanide and the nuclei around.[104,105] The effect is also distance-dependent, although decreases less rapidly with the distance than the NOE and the PRE (r^{-3} compared to r^{-6}). Therefore, it is possible to use PCS to obtain long-range distance constrains. $\Delta\delta_{PCS}$ is calculated as the difference in chemical shifts measured between diamagnetic and paramagnetic conditions.

$$\Delta\delta_{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}sin^2\theta cos2\varphi\right]$$

Equation 10. Equation of pseudo-contact shifts, dependent of the distance and the orientation.

Thus, the paramagnetic approach using PCS is a powerful technique to perform conformational analysis of carbohydrates. Simple disaccharide derivatives were first studied,[106-108] while more complex structures were then analyzed, breaking the limits of the NMR-based conformational elucidation of branched oligosaccharides.[109-112]

Moreover, the use of the lanthanide binding tags attached to the carbohydrates have opened new NMR-based avenues for exploring protein-glycan interactions, allowing the study of the molecular recognition features from two approaches: ligand-based and receptor-based. The receptor-based approach provides structural information of the binding event through the observation of PCSs in the protein NMR spectra, due to the paramagnetic effect induced by the lanthanide attached to the carbohydrate.[101]

On the other hand, the ligand-based approach does not depend on the receptor properties and may be applied to study a wide variety of bio macromolecules. Multiantenna N-glycans present an additional problem for NMR studies, besides the typical narrow signal dispersion. The isochronous chemical shifts presented for the same units located at different branches make impossible the distinction of their signals in the NMR spectra. With the introduction of a lanthanide binding tag, the generated PCSs expand the chemical shift dispersion into a much wider ppm range, allowing the unambiguous assignment of the carbohydrate NMR signals (figure 24). Thus, the conformational analysis of the glycan is possible, and more important, the exploration of the selectivity of the lectins for its different branches through line width analysis or in combination with STD NMR methods, as described in this Thesis.

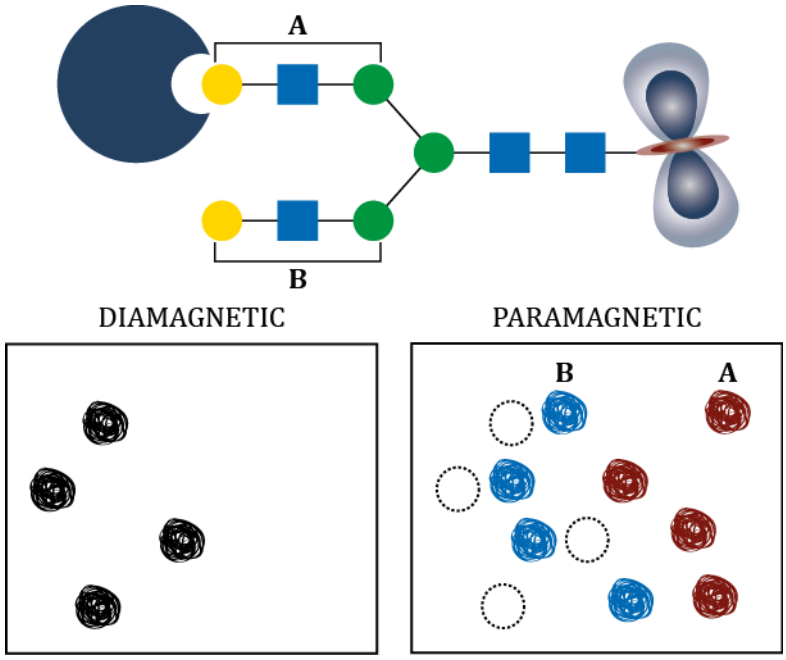


Fig 24. Schematic representation of the ^{1}H-^{13}C HSQC spectra of a N-glycan using the lanthanoid approach. The use of paramagnetic tags breaks the pseudosymmetry and different NMR signals can be observed for each residue at the different branches. Structural information and branch selectivity can now be deciphered.

1.4. REFERENCES

[1] C. R. Bertozzi, L. L. Kiessling, *Science.* **2001**, *291*, 2357–2364.
[2] "The official website of the Nobel Prize - NobelPrize.org," can be found under https://www.nobelprize.org/
[3] Y.-Y. Zhao, M. Takahashi, J.-G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume, N. Taniguchi, *Cancer Sci.* **2008**, *99*, 1304–1310.
[4] J. Ereño-Orbea, T. Sicard, H. Cui, M. T. Mazhab-Jafari, S. Benlekbir, A. Guarné, J. L. Rubinstein, J.-P. Julien, *Nat. Commun.* **2017**, *8*, 764.
[5] A. Varki, *Glycobiology* **2011**, *21*, 1121–1124.
[6] L. Sun, D. R. Middleton, P. L. Wantuch, A. Ozdilek, F. Y. Avci, *Glycobiology* **2016**, *26*, 1029–1040.
[7] G. A. Rabinovich, M. A. Toscano, *Nat. Rev. Immunol.* **2009**, *9*, 338–352.
[8] A. Varki, S. Kornfeld, *Historical Background and Overview*, Cold Spring Harbor Laboratory Press, **2015**.
[9] L. L. Kiessling, R. A. Splain, *Annu. Rev. Biochem.* **2010**, *79*, 619–653.
[10] E. Juaristi, G. Cuevas, *Tetrahedron* **1992**, *48*, 5019–5087.
[11] S. J. Angyal, *Angew. Chemie Int. Ed.* **1969**, *8*, 157–166.
[12] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
[13] L. Lebedel, A. Ardá, A. Martin, J. Désiré, A. Mingot, M. Aufiero, N. Aiguabella Font, R. Gilmour, J. Jiménez-Barbero, Y. Blériot, et al., *Angew. Chemie Int. Ed.* **2019**, *58*, 13758–13762.
[14] A. Martin, A. Arda, J. Désiré, A. Martin-Mingot, N. Probst, P. Sinaÿ, J. Jiménez-Barbero, S. Thibaudeau, Y. Blériot, *Nat. Chem.* **2016**, *8*, 186–191.
[15] H. B. Mayes, L. J. Broadbelt, G. T. Beckham, *J. Am. Chem. Soc.* **2014**, *136*, 1008–1022.
[16] J. Iglesias-Fernández, L. Raich, A. Ardèvol, C. Rovira, *Chem. Sci.* **2015**, *6*, 1167–1177.
[17] R. Kumar, B. Henrissat, P. M. Coutinho, *Sci. Rep.* **2019**, *9*, 10346.
[18] G. Speciale, A. J. Thompson, G. J. Davies, S. J. Williams, *Curr. Opin. Struct. Biol.* **2014**, *28*, 1–13.
[19] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, et al., *Glycobiology* **2015**, *25*, 1323–1324.
[20] P. H. Seeberger, *Monosaccharide Diversity*, Cold Spring Harbor Laboratory Press, **2015**.
[21] H. Rudiger, H.-C. Siebert, D. Solis, J. Jimenez-Barbero, A. Romero, C.-W. Lieth, T. Diaz-Maurino, H.-J. Gabius, *Curr. Med. Chem.* **2000**, *7*, 389–416.
[22] J. L. Asensio, F. J. Cañada, A. García-Herrero, M. T. Murillo, A. Fernández-Mayoralas, B. A. Johns, J. Kozak, Z. Zhu, C. R. Johnson, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **1999**, *121*, 11318–11329.
[23] P. Valverde, A. Ardá, N.-C. Reichardt, J. Jiménez-Barbero, A. Gimeno, *Medchemcomm* **2019**, *10*, 1678–1691.
[24] C. Wong, Ed. , *Carbohydrate-Based Drug Discovery*, Wiley, **2003**.
[25] S. Sattin, A. Bernardi, Royal Society Of Chemistry, **2015**, pp. 1–25.
[26] M. E. Cano, R. Agusti, A. J. Cagnoni, M. F. Tesoriero, J. Kovensky, M. L. Uhrig, R. M. de Lederkremer, *Beilstein J. Org. Chem.* **2014**, *10*, 3073–3086.

[27] E. Leclerc, X. Pannecoucke, M. Ethève-Quelquejeu, M. Sollogoub, *Chem. Soc. Rev.* **2013**, *42*, 4270–4283.
[28] P. Vidal, V. Roldós, M. del C. Fernández-Alonso, B. Vauzeilles, Y. Bleriot, F. J. Cañada, S. André, H.-J. Gabius, J. Jiménez-Barbero, J. F. Espinosa, et al., *Chem. - A Eur. J.* **2013**, *19*, 14581–14590.
[29] B. Xu, L. Unione, J. Sardinha, S. Wu, M. Ethève-Quelquejeu, A. Pilar Rauter, Y. Blériot, Y. Zhang, S. Martín-Santamaría, D. Díaz, et al., *Angew. Chemie Int. Ed.* **2014**, *53*, 9597–9602.
[30] I. Sutkeviciute, M. Thépaut, S. Sattin, A. Berzi, J. McGeagh, S. Grudinin, J. Weiser, A. Le Roy, J. J. Reina, J. Rojo, et al., *ACS Chem. Biol.* **2014**, *9*, 1377–1385.
[31] V. M. E. Andriotis, M. Rejzek, M. D. Rugen, B. Svensson, A. M. Smith, R. A. Field, *Biochem. Soc. Trans.* **2016**, *44*, 159–165.
[32] L. Unione, B. Xu, D. Díaz, S. Martín-Santamaría, A. Poveda, J. Sardinha, A. P. Rauter, Y. Blériot, Y. Zhang, F. J. Cañada, et al., *Chem. - A Eur. J.* **2015**, *21*, 10513–10521.
[33] N. Fontelle, A. Yamamoto, A. Arda, J. Jiménez-Barbero, A. Kato, J. Désiré, Y. Blériot, *Eur. J. Org. Chem.* **2018**, *2018*, 5477–5488.
[34] I. Calloni, L. Unione, G. Jiménez-Osés, F. Corzana, L. Del Bino, A. Corrado, O. Pitirollo, C. Colombo, L. Lay, R. Adamo, et al., *Eur. J. Org. Chem.* **2018**, *2018*, 4548–4555.
[35] H. K. Tam, J. Härle, S. Gerhardt, J. Rohr, G. Wang, J. S. Thorson, A. Bigot, M. Lutterbeck, W. Seiche, B. Breit, et al., *Angew. Chemie Int. Ed.* **2015**, *54*, 2811–2815.
[36] K. E. van Straaten, J. R. A. Kuttiyatveetil, C. M. Sevrain, S. A. Villaume, J. Jiménez-Barbero, B. Linclau, S. P. Vincent, D. A. R. Sanders, *J. Am. Chem. Soc.* **2015**, *137*, 1230–1244.
[37] J. D. Martínez, P. Valverde, S. Delgado, C. Romanò, B. Linclau, N. C. Reichardt, S. Oscarson, A. Ardá, J. Jiménez-Barbero, F. J. Cañada, *Molecules* **2019**, *24*, 2337.
[38] B. Ernst, J. L. Magnani, *Nat. Rev. Drug Discov.* **2009**, *8*, 661–677.
[39] A. Bernardi, P. Cheshev, *Chem. - A Eur. J.* **2008**, *14*, 7434–7441.
[40] P. Stanley, N. Taniguchi, M. Aebi, *N-Glycans*, Cold Spring Harbor Laboratory Press, **2015**.
[41] A. Varki, P. Gagneux, *Biological Functions of Glycans*, Cold Spring Harbor Laboratory Press, **2015**.
[42] A. Varki, *Glycobiology* **2017**, *27*, 3–49.
[43] W. J. Lockyer, in *Essentials ABO–Rh Group. Compat. Test.*, National Center For Biotechnology Information (US), **1982**, pp. 1–19.
[44] P. Rahfeld, L. Sim, H. Moon, I. Constantinescu, C. Morgan-Lang, S. J. Hallam, J. N. Kizhakkedathu, S. G. Withers, *Nat. Microbiol.* **2019**, *4*, 1475–1485.
[45] G. F. Clark, *Biochem. Biophys. Res. Commun.* **2014**, *450*, 1195–1203.
[46] A. B. Diekman, *Cell. Mol. Life Sci.* **2003**, *60*, 298–308.
[47] P.-C. Pang, P. C. N. Chiu, C.-L. Lee, L.-Y. Chang, M. Panico, H. R. Morris, S. M. Haslam, K.-H. Khoo, G. F. Clark, W. S. B. Yeung, et al., *Science.* **2011**, *333*, 1761–1764.
[48] S. P. Commins, H. R. James, L. A. Kelly, S. L. Pochan, L. J. Workman, M. S. Perzanowski, K. M. Kocan, J. V. Fahy, L. W. Nganga, E. Ronmark, et al., *J. Allergy Clin. Immunol.* **2011**, *127*, 1286-1293.

[49] R. J. Mullins, H. James, T. A. E. Platts-Mills, S. Commins, *J. Allergy Clin. Immunol.* **2012**, *129*, 1334-1342.
[50] A. J. Lutz, P. Li, J. L. Estrada, R. A. Sidner, R. K. Chihara, S. M. Downey, C. Burlak, Z.-Y. Wang, L. M. Reyes, B. Ivary, et al., *Xenotransplantation* **2013**, *20*, 27–35.
[51] D. K. C. Cooper, *Glycobiology* **2016**, *26*, 571–581.
[52] R. Dorsey, G. Emmett, H. Salem, *Handb. Toxicol. Chem. Warf. Agents* **2015**, 347–360.
[53] R. D. Cummings, R. L. Schnaar, *R-Type Lectins*, Cold Spring Harbor Laboratory Press, **2015**.
[54] "Poison umbrella murder case is reopened | The Independent," can be found under https://www.independent.co.uk/news/uk/crime/poison-umbrella-murder-case-is-reopened-851022.html
[55] J. L. Asensio, F. J. Canada, J. Jimenez-Barbero, *Eur. J. Biochem.* **1995**, *233*, 618–630.
[56] P. Fernández, J. Jiménez-Barbero, M. Martín-Lomas, D. Solís, T. Díaz-Mauriño, *Carbohydr. Res.* **1994**, *256*, 223–244.
[57] P. Fernández, J. Jiménez-Berbero, M. Martín-Lomas, *Carbohydr. Res.* **1994**, *254*, 61–79.
[58] A. Rivera-Sagredo, D. Solis, T. Diaz-Maurino, J. Jimenez-Barbero, M. Martin-Lomas, *Eur. J. Biochem.* **1991**, *197*, 217–228.
[59] J. Jiménez-Barbero, J. L. Asensio, F. J. Cañada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–55.
[60] H. Kogelberg, D. Solís, J. Jiménez-Barbero, *Curr. Opin. Struct. Biol.* **2003**, *13*, 646–53.
[61] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490–D495.
[62] G. Davies, B. Henrissat, *Structure* **1995**, *3*, 853–859.
[63] L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, *Annu. Rev. Biochem.* **2008**, *77*, 521–555.
[64] V. Lombard, T. Bernard, C. Rancurel, H. Brumer, P. M. Coutinho, B. Henrissat, *Biochem. J.* **2010**, *432*, 437–444.
[65] A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho, B. Henrissat, *Biotechnol. Biofuels* **2013**, *6*, 41.
[66] M. E. Taylor, K. Drickamer, R. L. Schnaar, M. E. Etzler, A. Varki, *Discovery and Classification of Glycan-Binding Proteins*, Cold Spring Harbor Laboratory Press, **2015**.
[67] H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, D. Solís, *Trends Biochem. Sci.* **2011**, *36*, 298–313.
[68] M. del Carmen Fernández-Alonso, F. J. Cañada, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* **2005**, *127*, 7379–86.
[69] K. Ramírez-Gualito, R. Alonso-Ríos, B. Quiroz-García, A. Rojas-Aguilar, D. Díaz, J. Jiménez-Barbero, G. Cuevas, *J. Am. Chem. Soc.* **2009**, *131*, 18129–18138.
[70] J. L. Asensio, A. Ardá, F. J. Cañada, J. Jiménez-Barbero, *Acc. Chem. Res.* **2013**, *46*, 946–54.
[71] L. Nieto, Á. Canales, G. Giménez-Gallego, P. M. Nieto, J. Jiménez-Barbero, *Chem. - A Eur. J.* **2011**, *17*, 11204–11209.
[72] M. del Carmen Fernandez-Alonso, D. Diaz, M. Alvaro Berbis, F. Marcelo, J. Canada, J. Jimenez-Barbero, *Curr. Protein Pept. Sci.* **2012**, *13*, 816–830.

[73] E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **1946**, *69*, 37–38.
[74] F. Bloch, W. W. Hansen, M. Packard, *Phys. Rev.* **1946**, *70*, 474–485.
[75] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, **2016**.
[76] H. (Harald) Günther, *NMR Spectroscopy : Basic Principles, Concepts, and Applications in Chemistry*, **2013**.
[77] M. del C. Fernández-Alonso, M. Alvaro Berbis, Á. Canales, A. Ardá, F. Javier Cañada, J. Jiménez-Barbero, **2013**, pp. 7–42.
[78] A. Ardá, A. Canales, F. J. Cañada, J. Jiménez-Barbero, Royal Society Of Chemistry, **2015**, pp. 1–20.
[79] M. Álvaro Berbís, Á. Canales, J. Sastre-Martinez, L. Unione, M. C. Fernández-Alonso, P. Blasco, F. Javier Cañada, J. Jiménez-Barbero, in *Carbohydr. Chem. State Art Challenges Drug Dev.*, Imperial College Press, **2015**, pp. 121–146.
[80] A. Ardá, H. Coelho, B. F. de Toro, S. Galante, A. Gimeno, A. Poveda, J. Sastre, L. Unione, P. Valverde, F. J. Cañada, et al., **2016**, pp. 47–82.
[81] A. Ardá, J. Jiménez-Barbero, *Chem. Commun.* **2018**, *54*, 4761–4769.
[82] P. Valverde, J. I. Quintana, J. I. Santos, A. Ardá, J. Jiménez-Barbero, *ACS Omega* **2019**, *4*, 13618–13630.
[83] A. Gimeno, P. Valverde, A. Ardá, J. Jiménez-Barbero, *Curr. Opin. Struct. Biol.* **2020**, *62*, 22–30.
[84] M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11–15.
[85] J. Ribeiro, T. Diercks, J. Jiménez-Barbero, S. André, H.-J. Gabius, F. Cañada, *Biomolecules* **2015**, *5*, 3177–3192.
[86] M. Delbianco, A. Kononov, A. Poveda, Y. Yu, T. Diercks, J. Jiménez-Barbero, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 5421–5426.
[87] D. Neuhaus, in *Encycl. Magn. Reson.*, John Wiley & Sons, Ltd, Chichester, UK, **2011**.
[88] B. Meyer, T. Peters, *Angew. Chemie Int. Ed.* **2003**, *42*, 864–890.
[89] M. Pellecchia, D. S. Sem, K. Wüthrich, *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219.
[90] M. Mayer, B. Meyer, *Angew. Chemie Int. Ed.* **1999**, *38*, 1784–1788.
[91] P. Kirsch, A. M. Hartman, A. K. H. Hirsch, M. Empting, *Molecules* **2019**, *24*, 4309.
[92] M. M. and, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.
[93] P. Balaram, A. A. Bothner-By, J. Dadok, *J. Am. Chem. Soc.* **1972**, *94*, 4015–4017.
[94] P. Balaram, A. A. Bothner-By, E. Breslow, *J. Am. Chem. Soc.* **1972**, *94*, 4017–4018.
[95] A. Poveda, J. Jiménez-Barbero, *Chem. Soc. Rev.* **1998**, *27*, 133.
[96] J. Jiménez-Barbero, T. Peters, in *NMR Spectrosc. Glycoconjugates*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, **2003**, pp. 289–310.
[97] M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
[98] N. Sibille, X. Hanoulle, F. Bonachera, D. Verdegem, I. Landrieu, J.-M. Wieruszeski, G. Lippens, *J. Biomol. NMR* **2009**, *43*, 219–227.
[99] C. Dalvit, N. Mongelli, G. Papeo, P. Giordano, M. Veronesi, D. Moskau, R. Kümmerle, **2005**, *127*, 13380-13385.
[100] N. U. Jain, A. Venot, K. Umemoto, H. Leffler, J. H. Prestegard, *Protein Sci.* **2008**, *10*, 2393–2400.

[101] Á. Canales, Á. Mallagaray, M. Á. Berbís, A. Navarro-Vázquez, G. Domínguez, F. J. Cañada, S. André, H.-J. Gabius, J. Pérez-Castells, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **2014**, *136*, 8011–8017.
[102] G. Otting, *J. Biomol. NMR* **2008**, *42*, 1–9.
[103] M. J. Moure, A. Eletsky, Q. Gao, L. C. Morris, J.-Y. Yang, D. Chapla, Y. Zhao, C. Zong, I. J. Amster, K. W. Moremen, et al., *ACS Chem. Biol.* **2018**, *13*, 2560–2567.
[104] I. Bertini, C. Luchinat, G. Parigi, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249–273.
[105] G. Parigi, E. Ravera, C. Luchinat, *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *114–115*, 211–236.
[106] M. Erdélyi, E. d'Auvergne, A. Navarro-Vázquez, A. Leonov, C. Griesinger, *Chem. - A Eur. J.* **2011**, *17*, 9368–9376.
[107] S. Yamamoto, T. Yamaguchi, M. Erdélyi, C. Griesinger, K. Kato, *Chem. - A Eur. J.* **2011**, *17*, 9280–9282.
[108] A. Mallagaray, A. Canales, G. Domínguez, J. Jiménez-Barbero, J. Pérez-Castells, *Chem. Commun.* **2011**, *47*, 7179.
[109] Y. Zhang, S. Yamamoto, T. Yamaguchi, K. Kato, *Molecules* **2012**, *17*, 6658–71.
[110] T. Yamaguchi, Y. Sakae, Y. Zhang, S. Yamamoto, Y. Okamoto, K. Kato, *Angew. Chemie Int. Ed.* **2014**, *53*, 10941–10944.
[111] A. Canales, A. Mallagaray, J. Pérez-Castells, I. Boos, C. Unverzagt, S. André, H.-J. Gabius, F. J. Cañada, J. Jiménez-Barbero, *Angew. Chemie Int. Ed.* **2013**, *52*, 13789–13793.
[112] A. Canales, I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, et al., *Angew. Chemie Int. Ed.* **2017**, *56*, 14987–14991.

CHAPTER 2

OBJECTIVES

The general scientific objective of this Thesis has been to gain knowledge and eventually to understand, at the maximum possible level, the molecular basis of the interactions of carbohydrates and glycomimetics with their biological receptors with two specific systems:

-The interaction of novel *E. coli* β-galactosidase inhibitors with this particular enzyme.

-The interaction of complex type N-glycans with different Hemagglutinin strains of the Influenza A virus.

From the methodological perspective, the aim of this Thesis has been to develop novel methodologies, based on paramagnetic NMR, to elucidate the conformations, dynamics and interactions of complex type N-glycans.

Additionally, from the training perspective, the key objective has been to acquire demonstrated knowledge in biological chemistry, essentially in the design of protocols to unravel structural, conformational, dynamic, and molecular recognition features of systems of biological interest by applying different Nuclear Magnetic Resonance spectroscopy methodologies in combination with other techniques.

CHAPTER 3

DIASTEREOMERIC GLYCOSYL SULFOXIDES: CONFORMATIONAL ANALYSIS AND RECOGNITION FEATURES VERSUS *E. COLI* β-GALACTOSIDASE.

3.1. INTRODUCTION

Thiooligosaccharides are one of the most common types of glycomimetics. They have shown interesting biological properties and have been useful for studying different events, including metabolic processes.[1] The replacement of the oxygen atom of the interglycosidic linkage by a sulfur atom confers stability against the hydrolysis by glycosidases. In this context, they have been shown to act as moderate inhibitors of these enzymes.[2-5]

Thiodisaccharide sulfoxides are usually prepared by oxidizing the sulfur atom of the interglycosidic linkage. This results in two possible enantiomers. The configuration of the sulfur atom is important since the behavior may depend on the stereochemistry. For example, well known drugs, such as esomeprazole (*S* enantiomer)[6] have been selected from the racemic mixture to be used as pharmaceuticals. Furthermore, it has been reported that diastereomeric carbohydrate sulfoxides are enzymatically hydrolyzed in a diastereoselective manner.[7]

Therefore, the determination of the conformation and interaction properties of these molecules is of relevance. In this context, the work described in the following sections of this chapter is focused on the study of four sulfoxide disaccharides.

3.2. RESULTS AND DISCUSSION

The work presented herein has been carried out with a set of sulfoxide disaccharides (**1*S***, **1*R***, **2*S***, **2*R***) synthesized by the group of Dr. Oscar Varela at the University of Buenos Aires.[8,9]

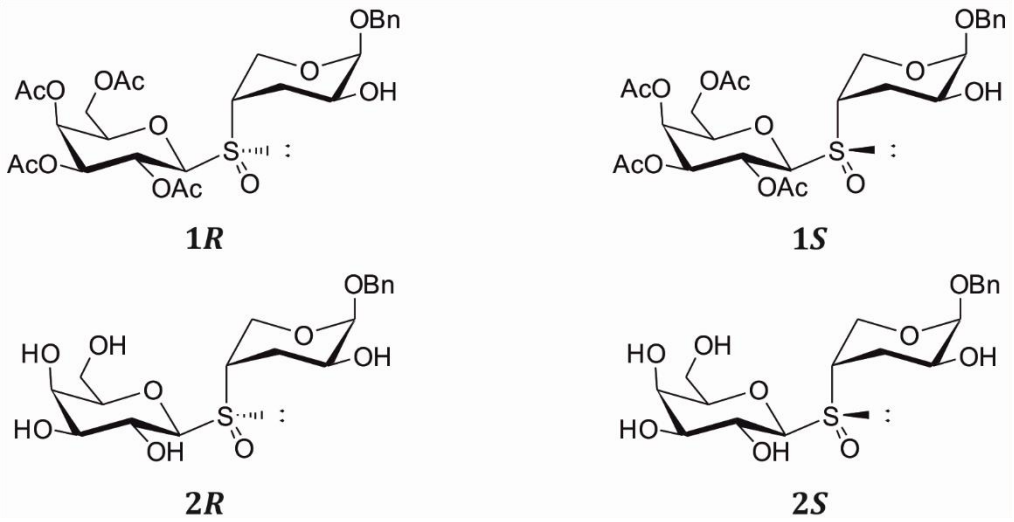


Fig 1. Structure of the sulfoxide disaccharides.

3.2.1. Conformational analysis of 1*R* and 1*S*

First, the conformational analysis of the acetylated derivatives (1*R* and 1*S*) was carried out. The analysis of the experimentally determined vicinal couplings values of the aglycon moiety, for both 1*R* and 1*S*, suggested the presence of an equilibrium between the $^{1}C_{4}$ and $^{4}C_{1}$ geometries (figure 2), whereas for the galactose moiety only the $^{4}C_{1}$ geometry was deduced. The contribution of each chair was estimated by comparing the experimental coupling constants with those calculated according to the generalized Karplus equation proposed by Altona, as implemented in the MSpin program (table 1).[10]

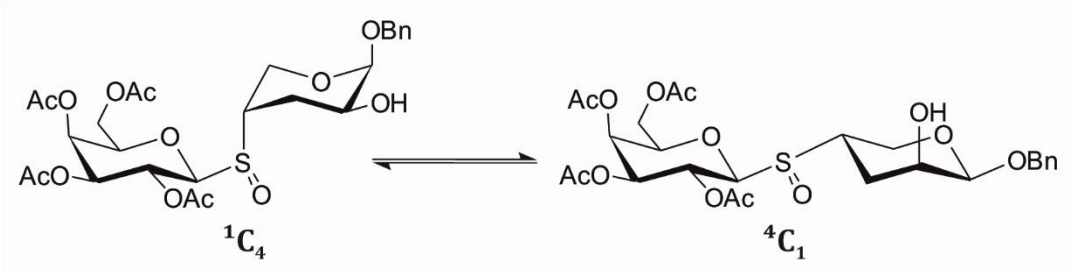


Fig 2. Conformational equilibrium of the aglyconic ring of **1*R*** and **1*S***.

^3J(H-H)	**1*R***			**1*S***		
	*J*calc. (Hz)		*J*exp. (Hz)	*J*calc. (Hz)		*J*exp. (Hz)
	$^1C_4$	$^4C_1$		$^1C_4$	$^4C_1$	
H_1-H_2	3.3	1.5	3.2	3.1	1.6	2.6
H_2-H_{3A}	11.1	3.4	10.2	11.1	3.2	8.9
H_2-H_{3B}	4.6	2.7	4.2	4.6	2.9	ca. 5.3
H_{3A}-H_4	4.7	3.7	4.7	5.2	3.8	4.6
H_{3B}-H_4	2.1	12.3	4.2	1.8	12.3	ca. 5.3
H_4-H_{5A}	3.1	4.5	3.1	3.7	4.3	ca. 4.0
H_4-H_{5B}	1.4	11.6	3.3	1.2	11.6	ca. 4.0

Table 1. Experimental and calculated $^3J_{HH}$ (Hz) of the aglycon ring of **1*R*** and **1*S***.

The data indicate that the $^1C_4$ conformer is the major form in both cases, contributing to the conformational equilibrium with 84% and 64% for **1*R*** and **1*S*** respectively.

The existence of a conformational equilibrium between the $^1C_4$ and $^4C_1$ geometries was confirmed by the characteristic Nuclear Overhauser Effect (NOE) cross-peaks observed in the 2D NOESY spectrum (figure 3). The observation of NOE cross-peaks between H_{3A}-H_{5A} assessed the presence of the major $^1C_4$ geometry, while the NOE cross-peaks between H_1-H_{3B} and H_1-H_{5B} noted the minor existence of the $^4C_1$ geometry (figure 4).

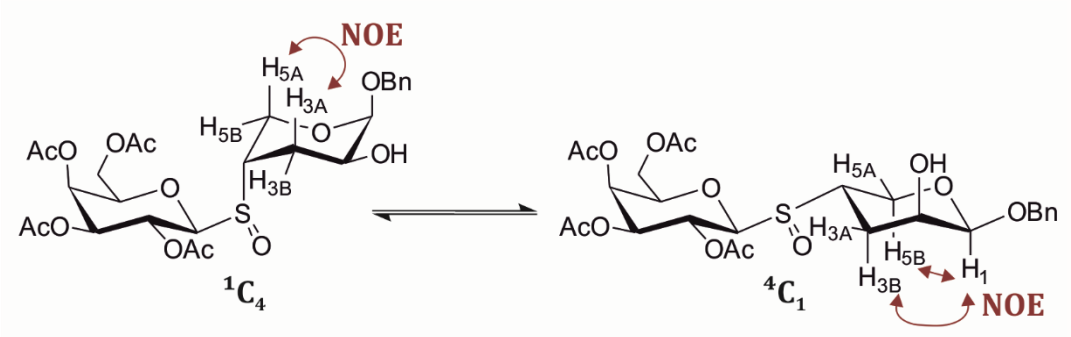


Fig 3. Intra-residue NOE cross peaks expected for each geometry of the conformational equilibrium of the aglyconic ring of **1*R*** and **1*S***.

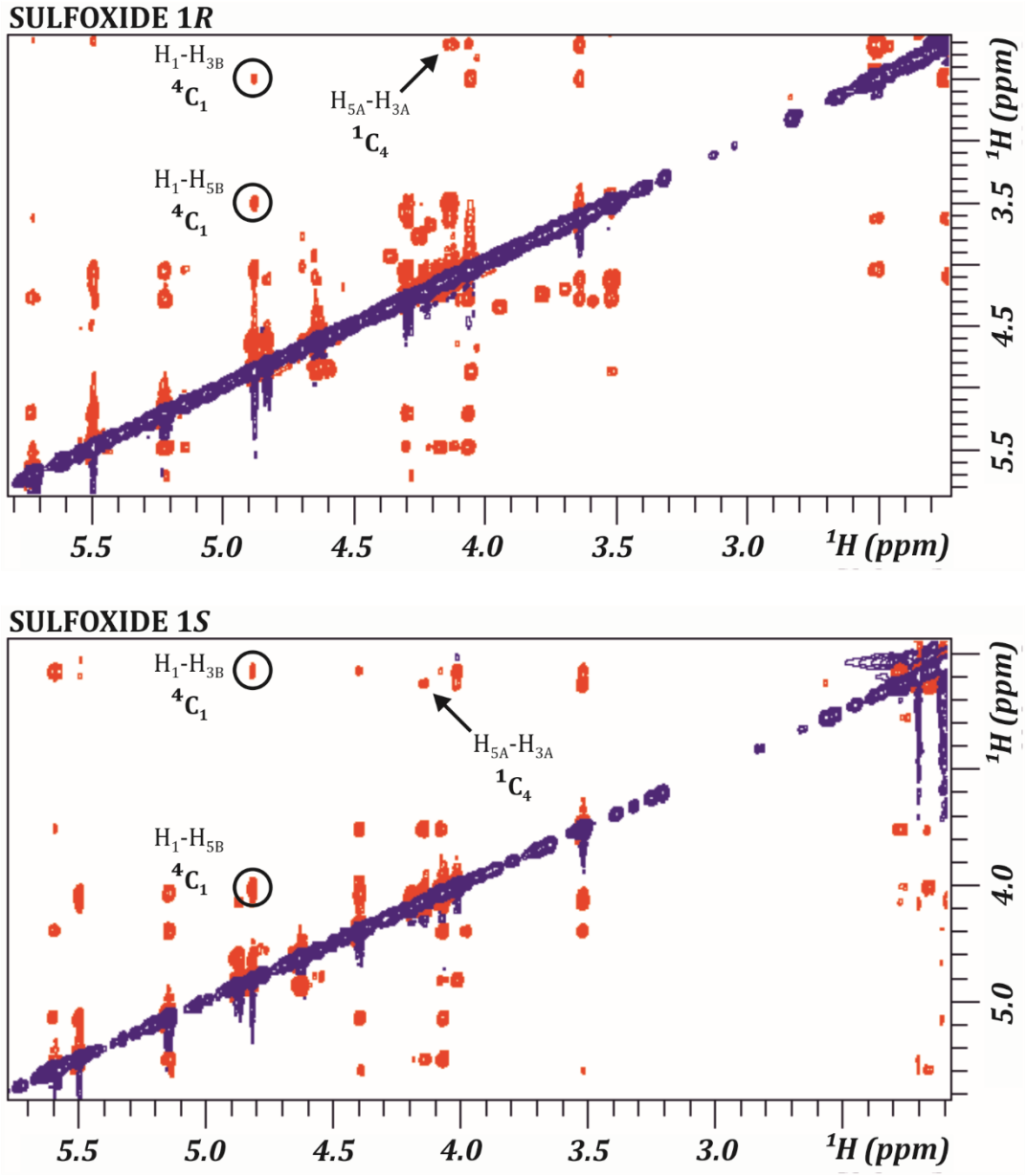


Fig 4. 2D-NOESY spectra recorded for **1*R*** (top) and **1*S*** (bottom). Key NOE cross-peaks for the $^{1}C_{4}$ and $^{4}C_{1}$ chairs are highlighted by arrows and circles respectively.

The next step was to elucidate the preferred conformation around the glycosidic sulfoxide linkage. Different geometries can be generated by rotation of the torsion angles defined as $\Phi = H_{1'}\text{-}C_{1'}\text{-}S\text{-}C_{4}$ and $\Psi = C_{1'}\text{-}S\text{-}C_{4}\text{-}H_{4}$ (figure 5).

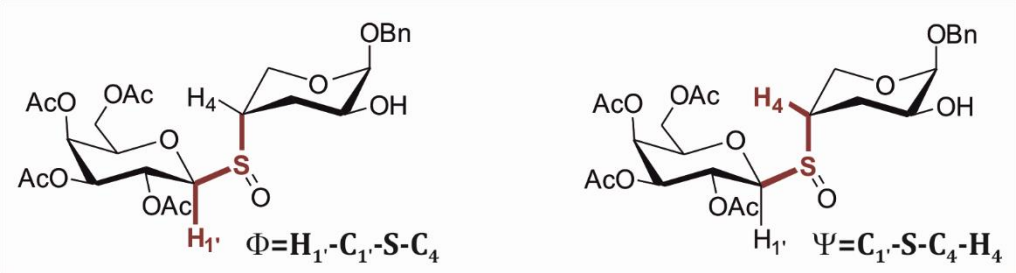


Fig 5. The torsion angles around the glycosidic sulfoxide linkage that define the different conformers.

The presence of a given rotamer in the conformational equilibrium was elucidated by analyzing the characteristic NOE cross-peaks observed in the 2D NOESY spectrum. This methodology has been previously employed to perform the conformational analysis of thiodisaccharides.[5] For the conformational analysis, the most populated $^1C_4$ conformer was first considered. The intense NOE interactions between $H_{1'}$-H_4 and $H_{1'}$-H_{5B} observed in the NOESY spectrum of **1*R*** indicated the existence of a highly populated *syn*-Φ/*syn*-Ψ conformer. Moreover, a weak NOE contact between $H_{2'}$-H_4, suggested the minor presence of the *anti*-Φ/*syn*-Ψ conformation (figure 6).

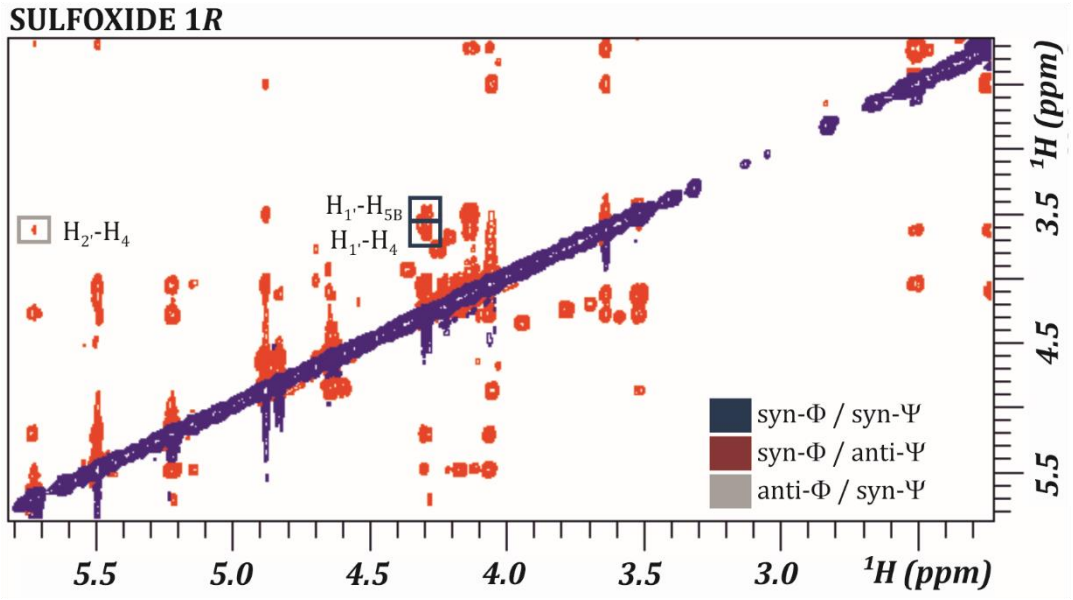


Fig 6. 2D-NOESY spectra recorded for **1*R***. Key NOE cross-peaks for the *syn*-Φ/*syn*-Ψ and *anti*-Φ/*syn*-Ψ conformations are highlighted in the blue and grey boxes respectively.

The same analysis was performed for **1*S***. An intense NOE interaction between $H_{1'}$-H_4 was observed, pointing out a highly populated *syn*-Φ/*syn*-Ψ conformer, as in the case of **1*R***. However, in this case, the minor *anti*-Φ/*syn*-Ψ conformation deduced by the interaction between $H_{2'}$-H_{3B} and $H_{2'}$-H_4 was accompanied by another minor geometry, *syn*-Φ/*anti*-Ψ, encoded by the NOE interaction between $H_{1'}$-H_{3B} (figure 7).

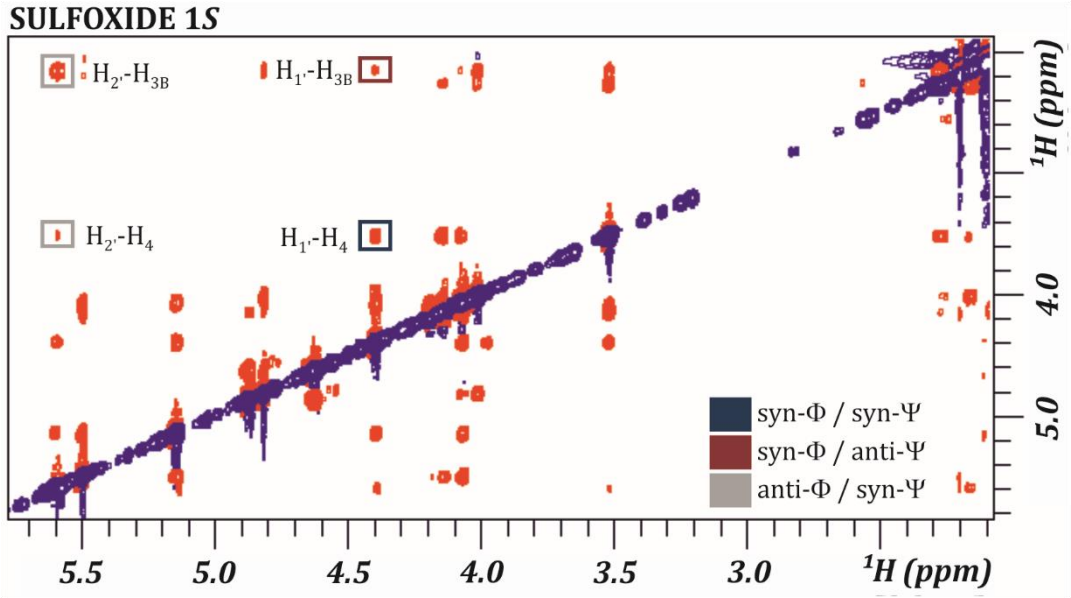


Fig 7. 2D-NOESY spectra recorded for **1*S***. Key NOE cross-peaks for the *syn*-Φ/*syn*-Ψ, *syn*-Φ/*anti*-Ψand *anti*-Φ/*syn*-Ψ conformations are highlighted in the blue, maroon and grey boxes respectively.

Once the conformational analysis of **1*R*** and **1*S*** had revealed a preference for the $^1C_4$ conformer for the aglycon and a different behavior around the glycosidic sulfoxide linkage for the different enantiomers, the deacetylation reaction of **1*R*** and **1*S*** was performed. Thus, the diastereomeric glycosyl sulfoxides **2*S*** and **2*R*** were obtained.

3.2.2. <u>Conformational analysis of 2R and 2S</u>

The conformational analysis of **2*R*** and **2*S*** was conducted in the same manner as for **1*R*** and **1*S***. First, ^1H-^1H coupling constants (*J*) were extracted from the 1D ^1H NMR spectra for both sulfoxides. As previously stated for the analogous acetylated thioglycoside derivatives, the experimental *J* values of the aglycon ring for both **2*R*** and **2*S*** compounds were indicative of the existence of a conformational equilibrium between the $^1C_4$ and $^4C_1$ geometries (figure 8).

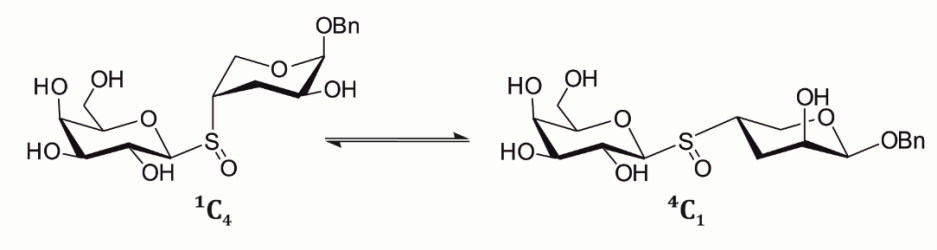


Fig 8. Conformational equilibrium of the aglyconic ring of **2*R*** and **2*S***.

The contribution of each chair was estimated by comparing the experimental *J* values and those calculated following the same protocol used before (table 2).

	2*R*			**2*S***		
^3J(H-H)	***J*calc. (Hz)**		***J*exp. (Hz)**	***J*calc. (Hz)**		***J*exp. (Hz)**
	$^1C_4$	$^4C_1$		$^1C_4$	$^4C_1$	
H_1-H_2	3.3	1.5	3.3	3.1	1.5	3.4
H_2-H_{3A}	11.1	3.3	10.7	11.2	3.2	11.2
H_2-H_{3B}	4.6	2.9	5.2	4.6	2.9	3.8
H_{3A}-H_4	4.8	3.8	3.8	5.1	3.9	4.8
H_{3B}-H_4	2.0	12.3	3.8	1.8	12.2	4.9
H_4-H_{5A}	3.1	4.6	3.8	3.7	4.4	2.8
H_4-H_{5B}	1.4	11.6	3.8	1.1	11.6	---

Table 2. Experimental and calculated coupling constant values (Hz) of the aglycon ring of **2*R*** and **2*S***.

In this case, the analysis of the data indicated that the $^1C_4$ conformer is again the major form in both cases (75%), also supported by the analysis of the intra-residue cross-peaks found in the 2D-NOESY spectra. The analogous analysis for the galactose residue suggested that the typical $^4C_1$ geometry was the exclusive conformation for this ring (figures 9 and 10).

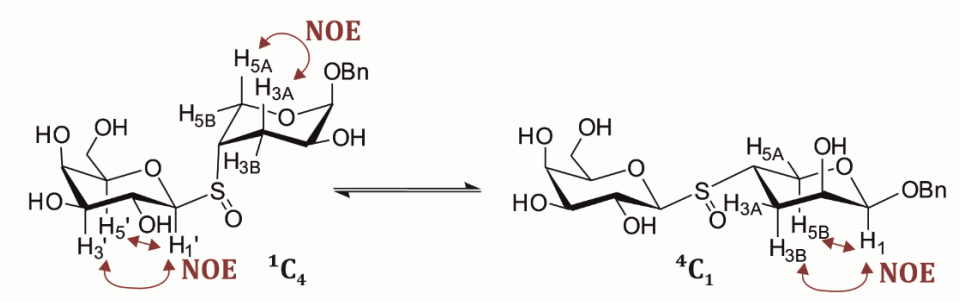


Fig 9. Intra-residue NOE cross peaks expected for each geometry of the conformational equilibrium of the aglyconic ring of **2*R*** and **2*S***, and exclusive cross peaks of the $^4C_1$ geometry of the galactose residue.

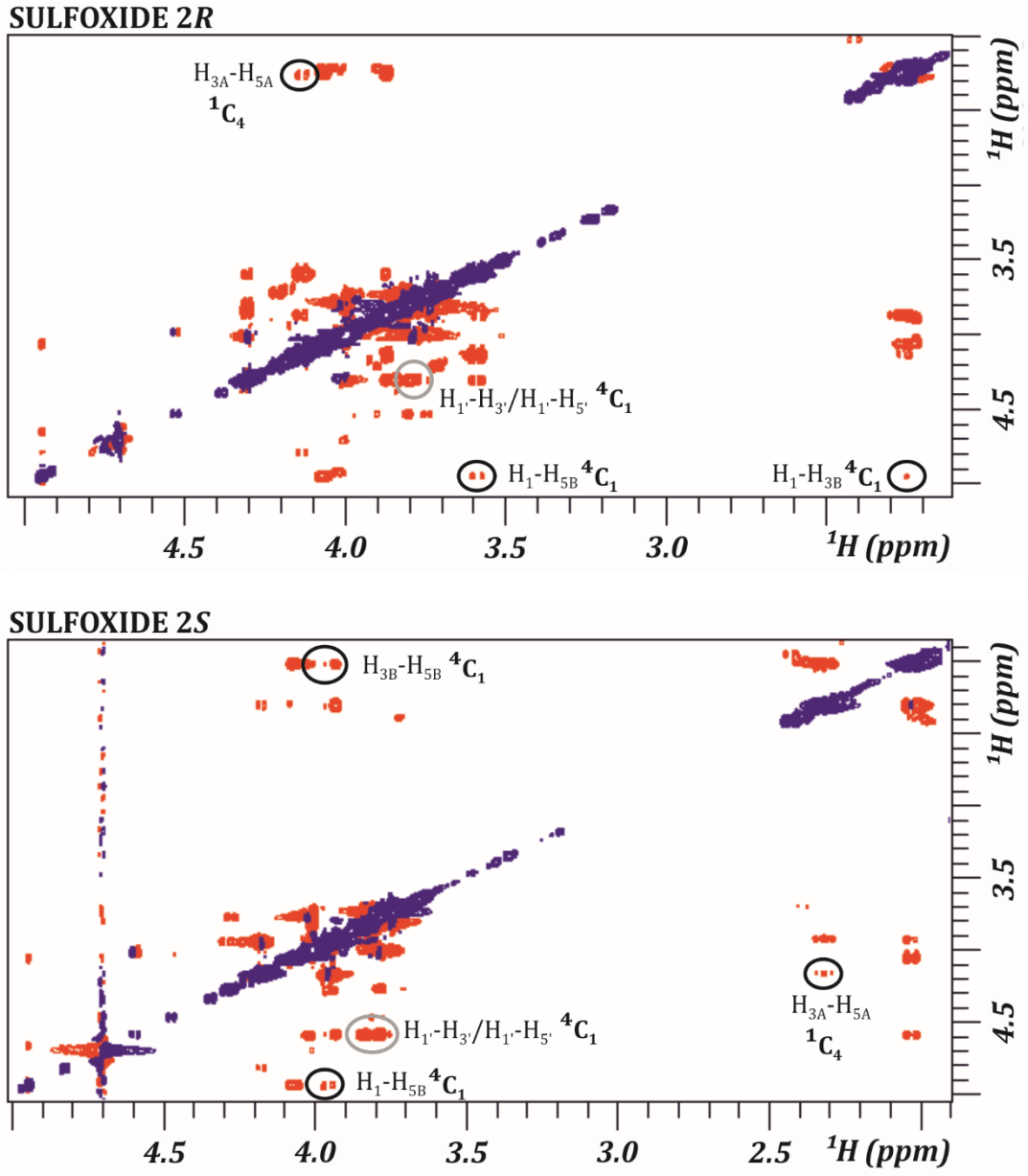


Fig 10. 2D-NOESY spectra recorded for **2*R*** (top) and **2*S*** (bottom). Key NOE cross-peaks for the $^{1}C_4$ and $^{4}C_1$ chairs of the aglycon ring are highlighted in black circles. Key NOE cross-peaks for the $^{4}C_1$ chair of the galactose residue are highlighted in grey circles.

The study of the conformation around the glycosidic sulfoxide linkage was subsequently performed. The torsion angles are defined as $\Phi = H_{1'}\text{-}C_{1'}\text{-}S\text{-}C_4$ and $\Psi = C_{1'}\text{-}S\text{-}C_4\text{-}H_4$, and the major contribution to the conformational equilibrium was elucidated by analyzing the NOE cross-peaks, as before. In the case of **2*R***, the observed NOE between $H_{1'}$-H_4 and the lack of NOE cross-peaks between $H_{2'}$-H_4 and $H_{1'}$-H_3 strongly suggested the existence of an exclusive *syn*-Φ/ *syn*-Ψ conformation (figure 11).

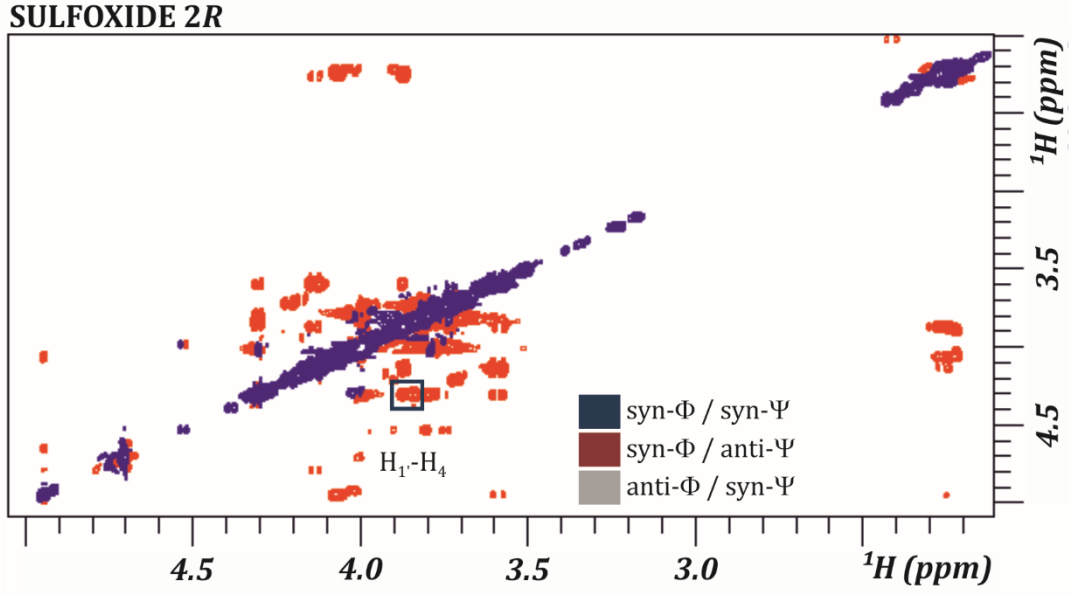


Fig 11. 2D-NOESY spectra recorded for **2*R***. Key NOE cross-peak for the *syn*-Φ/*syn*-Ψ conformation is highlighted in a blue box.

For **2*S***, an intense NOE was observed between $H_{1'}$-H_4, indicating that the *syn*-Φ/ *syn*-Ψ conformation is the most populated geometry. However, in this case, two other minor conformers were present in solution, the *syn*-Φ/*anti*-Ψ and *anti*-Φ/*syn*-Ψ conformations, encoded by the presence of $H_{1'}$-H_{3B} and $H_{2'}$-H_4 (weak) NOE cross-peaks, respectively (figure 12).

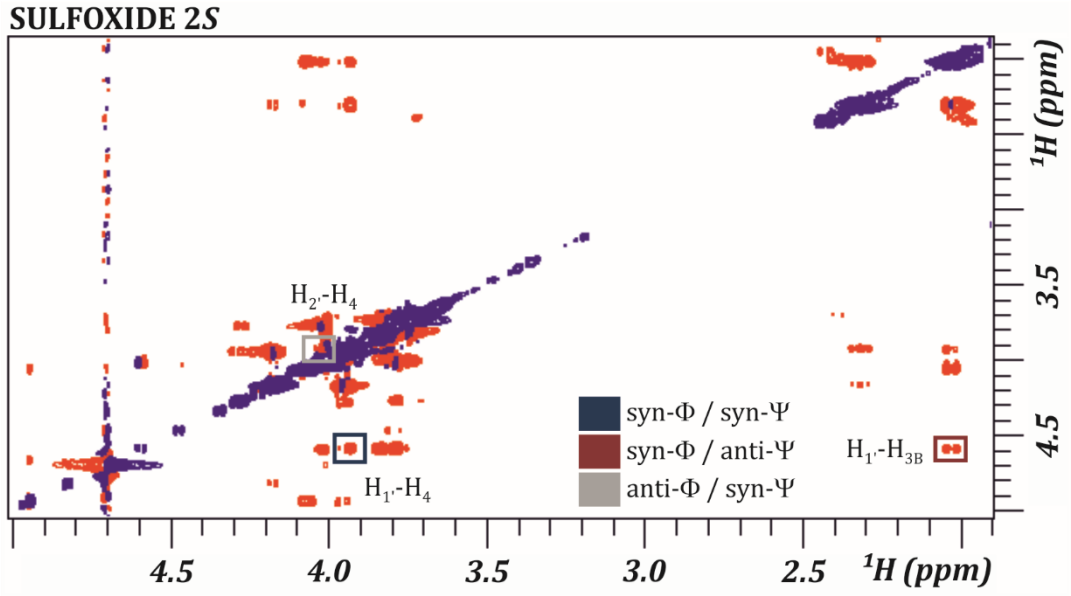


Fig 12. 2D-NOESY spectra recorded for **2*S***. Key NOE cross-peaks for the *syn*-Φ/*syn*-Ψ, *syn*-Φ/*anti*-Ψand *anti*-Φ/*syn*-Ψ conformations are highlighted in blue, maroon and grey boxes respectively.

3.2.3. Ab initio and molecular dynamics (MD) simulations for 2R and 2S

Therefore, the difference in the configuration at the stereogenic centre between **2*R*** and **2*S*** affects the conformational behavior of both sulfoxides. As stated above, only one conformation was found for **2*R*** (*syn*-Φ/*syn*-Ψ conformation), while three conformations were observed for **2*S*** (*syn*-Φ/*syn*-Ψ, *syn*-Φ/*anti*-Ψ and *anti*-Φ/*syn*-Ψ).

In order to unravel these behaviour differences, computational chemistry calculations using *ab initio* methods combined with molecular dynamics (MD) simulations were performed.

The stability difference between the three conformations of **2*S*** was obtained by quantum chemistry calculations. According to the employed method, the energy difference between the most abundant geometry, the *syn*-Φ/*syn*-Ψ, to the *anti*-Φ/*syn*-Ψ conformation was 1.8 kcal/mol, while that between the *anti*-Φ/*syn*-Ψ to *syn*-Φ/*anti*-Ψ amounted to 2 kcal/mol. In contrast, for **2*R*** the relative energy difference between the global minimum and the "theoretical" *anti*-Φ/*syn*-Ψ geometry was estimated as 3 kcal/mol, significantly higher than that for **2*S***. Fittingly, only the global minimum, the *syn*-Φ/*syn*-Ψ conformation was experimentally observed.

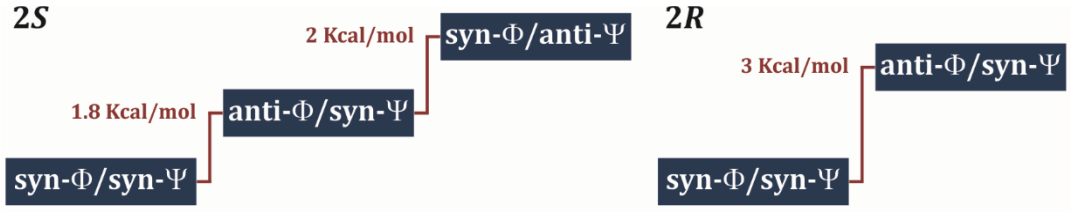


Fig 12. Stability difference between conformations for **2*S*** (left) and **2*R*** (right).

Thus, the calculated data were in full agreement with the presence of the *anti*-Φ/*syn*-Ψ conformation in solution for **2*S***, absent in **2*R***.

MD simulations for both compounds in the free state were then carried out. For **2*R***, the *syn*-Φ/*syn*-Ψ conformer was used as starting geometry, since it is the only conformation in solution. On the other hand, three MD simulations were computed for **2*S***, using as starting geometries the three conformations found in solution.

For **2*R***, the starting geometry (*syn*-Φ/*syn*-Ψ) remained stable along the MD trajectory (figure 13).

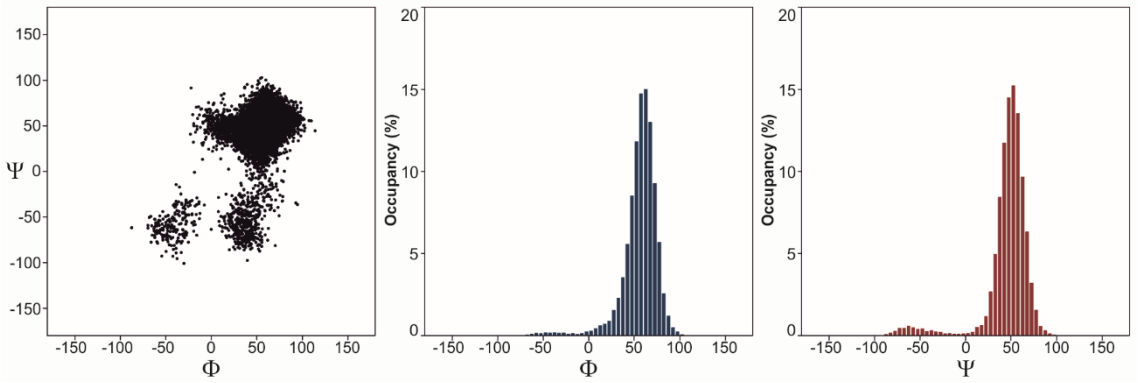


Fig 13. Left: Plot of the accessible Φ/Ψ torsion angles around the sulfoxide linkage during the MD simulation of **2*R***. **Right:** Distribution of Φ (blue) and Ψ (maroon) angles during the same simulation.

In contrast, the three MD simulations carried out for **2*S*** showed conformational changes during the calculations, independently of the starting geometry. Transitions between the *syn*-Φ/*syn*-Ψ, *syn*-Φ/*anti*-Ψ and *anti*-Φ/*syn*-Ψ conformers were always detected (figure 14).

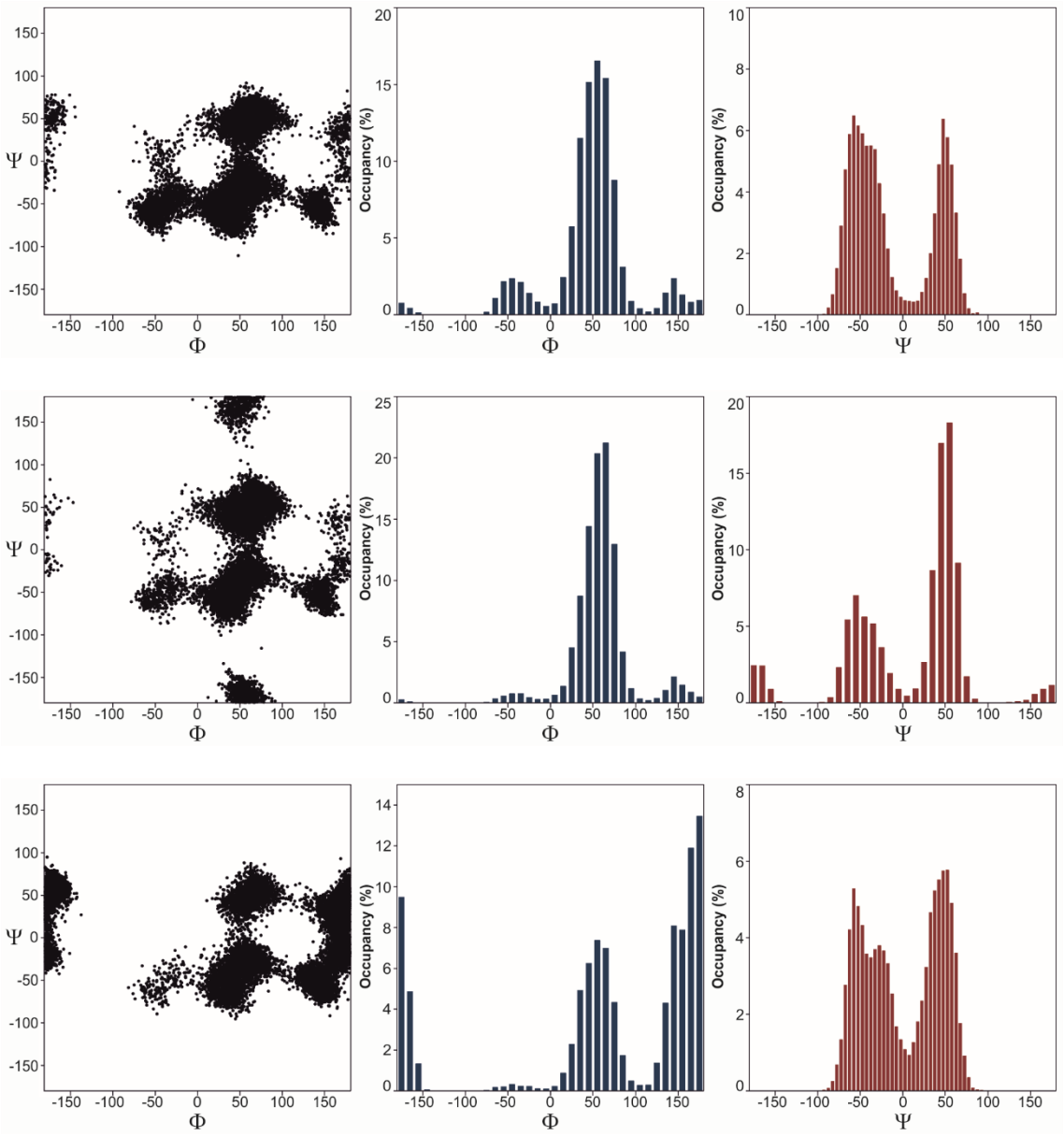


Fig 14. Left: Plots of the Φ/Ψ torsion angles accessible around the sulfoxide linkage of **2*S*** during the MD simulations. **Right:** Distribution of Φ (blue) and Ψ (maroon) angles during the same simulations. The starting geometries are *syn*-Φ/*syn*-Ψ (top), *syn*-Φ/*anti*-Ψ (centre) and *anti*-Φ/*syn*-Ψ (bottom).

Thus, these computational results are in agreement with the experimental observations that demonstrated that sulfoxide **2*S*** is more flexible than **2*R***.

3.2.4. The interaction of 2R and 2S with E. coli β-galactosidase

The interaction between both ligands and the *E. coli* β-galactosidase was analyzed. The acquisition of ^{1}H NMR spectra at different times demonstrated that, despite being a moderate inhibitor, a slow hydrolysis process exists for **2*R*** by the enzyme. After 2h 30 min of incubation with the β-galactosidase, a small amount of galactose wass detected (the hydrolysis product of **2*R***). However, a significant amount of **2*R*** sulfoxide remained unaffected after 20h of incubation (figure 15).

The same analysis was performed for **2*S*** sulfoxide and, contrary to **2*R***, a very fast hydrolysis was observed in the presence of the β-galactosidase (figure 16).

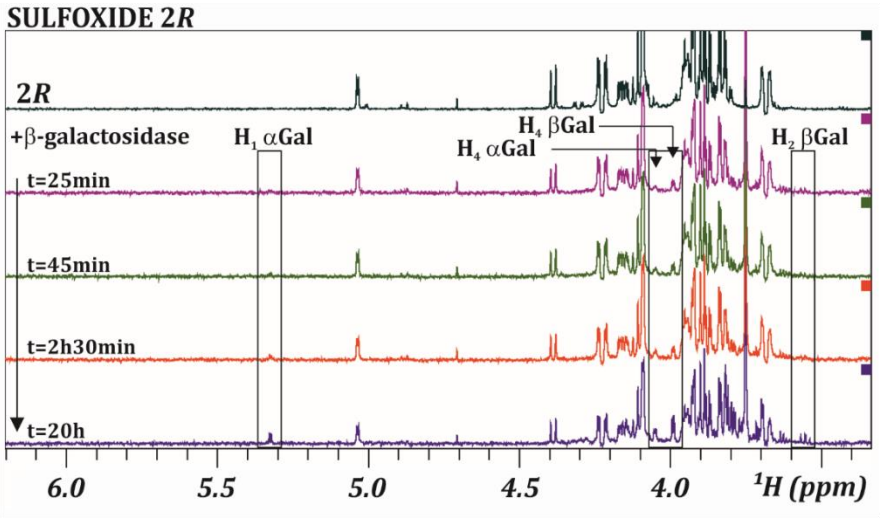


Fig 15. ^{1}H NMR spectrum of **2*R*** (top) and in the presence of *E. coli* β-galactosidase at different times. The evolution of the released α/β-galactose is highlighted.

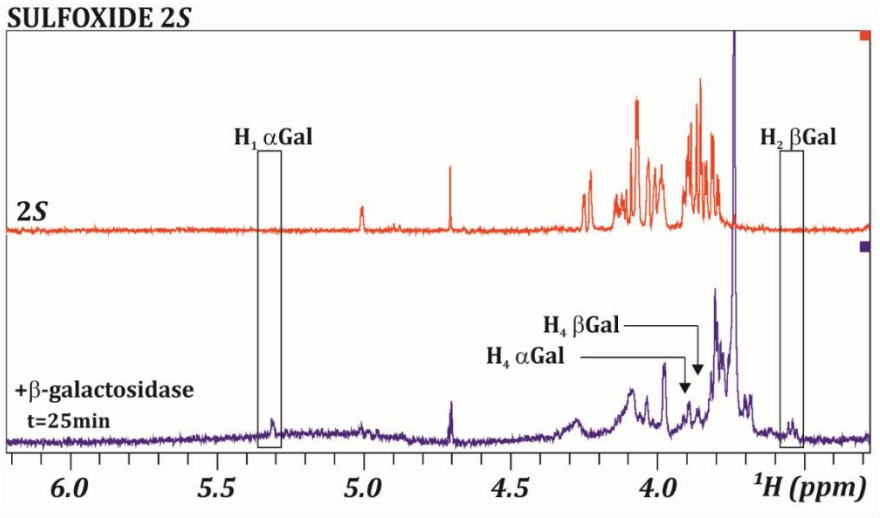


Fig 16. ^{1}H NMR spectrum of **2*S*** (top) and in the presence of *E. coli* β-galactosidase after 25 minutes. The presence of the released α/β-galactose is highlighted.

Since the hydrolysis observed of **2*R*** is slow, the key features of the molecular recognition process can be addressed using ligand-based STD NMR experiment. The STD signals for **2*R*** in the presence of the enzyme pointed out that the protons with the highest STD intensities belong to the galactose unit, specifically H_2, H_3 and H_4 (figure 17). Minor STD was observed for the aglycon.

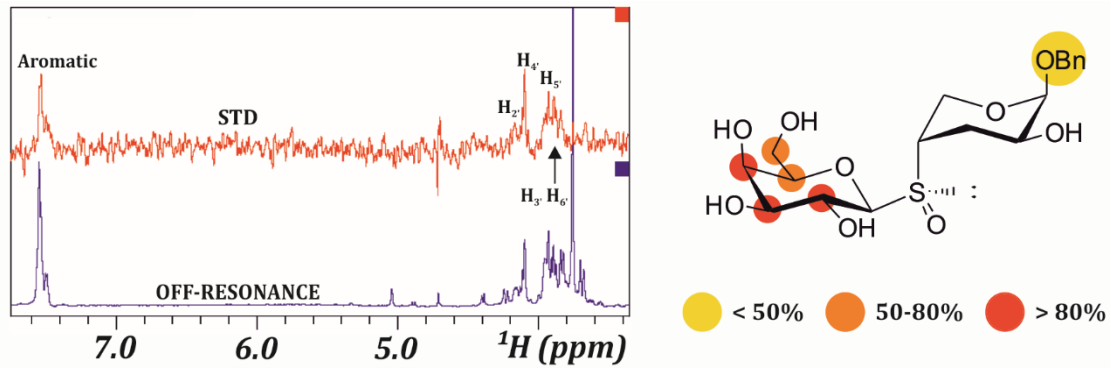
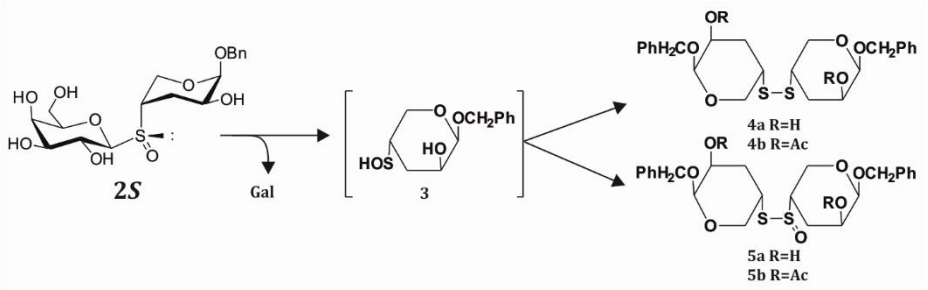


Fig 17. Left: STD NMR experiment of **2*R***. Signals observed in the STD spectrum belong to the protons that are involved in the recognition process. **Right:** Schematic representation of the observed STD percentages of 2R. The highest STD intensity protons belong to the galactose unit.

As previously shown, the hydrolysis of **1*S*** sulfoxide occurs rapidly. Galactose and the aglycon ring are expected to be released when the hydrolysis process takes place. However, the formation of an insoluble material was observed during the hydrolysis reaction. To investigate the outcome of this fragment, the group of Dr. Oscar Varela isolated the material and subjected it to acetylation for chromatographic purification. Two fractions were obtained, whose main components were identified by HRMS (ESI) (**4b** and **5b** products Scheme 1). The formation of these products can be justified if the hydrolysis of **2*S*** by the β-galactosidase produces the sulfenic acid **3** as a reactive intermediate. Due to their instability, sulfenic acids experience redox processes which lead to varied products (Scheme 1).



Scheme 1: Reaction scheme of **2*S*** sulfoxide.

3D models of **2*R*** and **2*S*** in the binding site of the enzyme were then obtained. The template employed was the geometry deduced in the elucidation of the X-ray structure of *E. coli* β-galactosidase in complex with galactose (PDB 1JZ7). It has been previously described that the high-energy *anti*-Φ/*syn*-Ψ conformer of C- and S-glycomimetics is recognized by the *E. coli* β-galactosidase through a conformational selection process.[5,11]

Indeed, the binding mode obtained by docking the *anti*-Φ/*syn*-Ψ conformers for both **2*R*** and **2*S*** sulfoxides in the binding site, satisfactorily explained the hydrolysis process. The amino acid responsible of the nucleophilic attack, Glu^{537}, is rather close to the anomeric position. Furthermore, the hydroxy group at C_2 of the galactose residue, which is involved in the recruitment of Glu^{537}, displays the proper orientation for the hydrolysis process.[12] In addition, the hydroxy group of Glu^{461}, which acts as the proton donor in the enzymatic process,[13] is oriented towards the sulfoxide group, close to the oxygen atom (figure 18). The beginning of the hydrolysis could be favored by this disposition, permitting the protonation of the oxygen atom of the sulfoxide, in agreement with the mechanism that involves a sulfenic acid as intermediate.

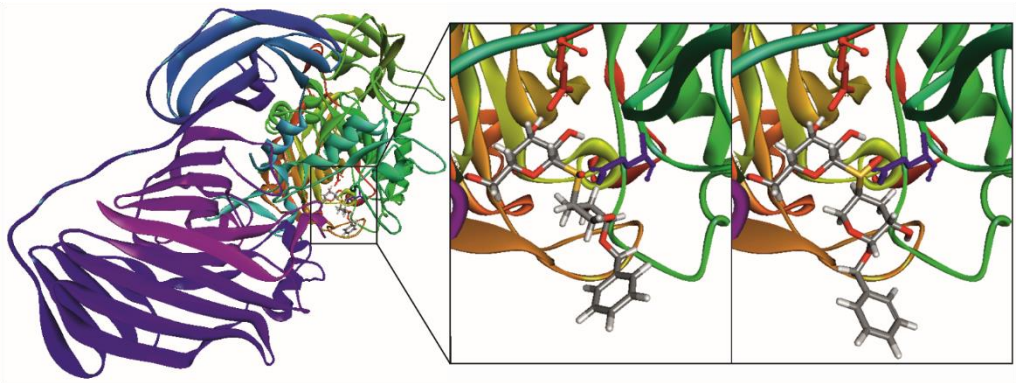


Fig 18. Depiction of the complex of one monomer domain of *E. coli* β-galactosidase docked with the *anti*-Φ/*syn*-Ψ conformer of **2*R*** (right) and **2*S*** (left). The proper disposition of the amino acids for the initial hydrolysis process in both cases can be observed. The hydroxy group of the proton donor, Glu^{461} (in blue) is oriented towards the oxygen atom of the sulfoxide group and the amino acid responsible of the nucleophilic attack, Glu^{537} (in red), is properly oriented.

The comparison of the docked *anti*-Φ/*syn*-Ψ and *syn*-Φ/*syn*-Ψ conformers for both sulfoxides in the binding site reveals that only the *anti*-Φ/*syn*-Ψ conformation suitably fits and displays the preferential geometry for the aglycon to behave as

leaving group. According to the NMR data, the *anti*-Φ/*syn*-Ψ conformation is only found in solution for **2*S***. Moreover, the *syn*-Φ/*syn*-Ψ conformation presents heavy steric clashes with the surrounding amino acid residues (figure 19). In this manner, the occurrence of the hydrolysis process, preferentially in the **2*S*** isomer, could be explained.

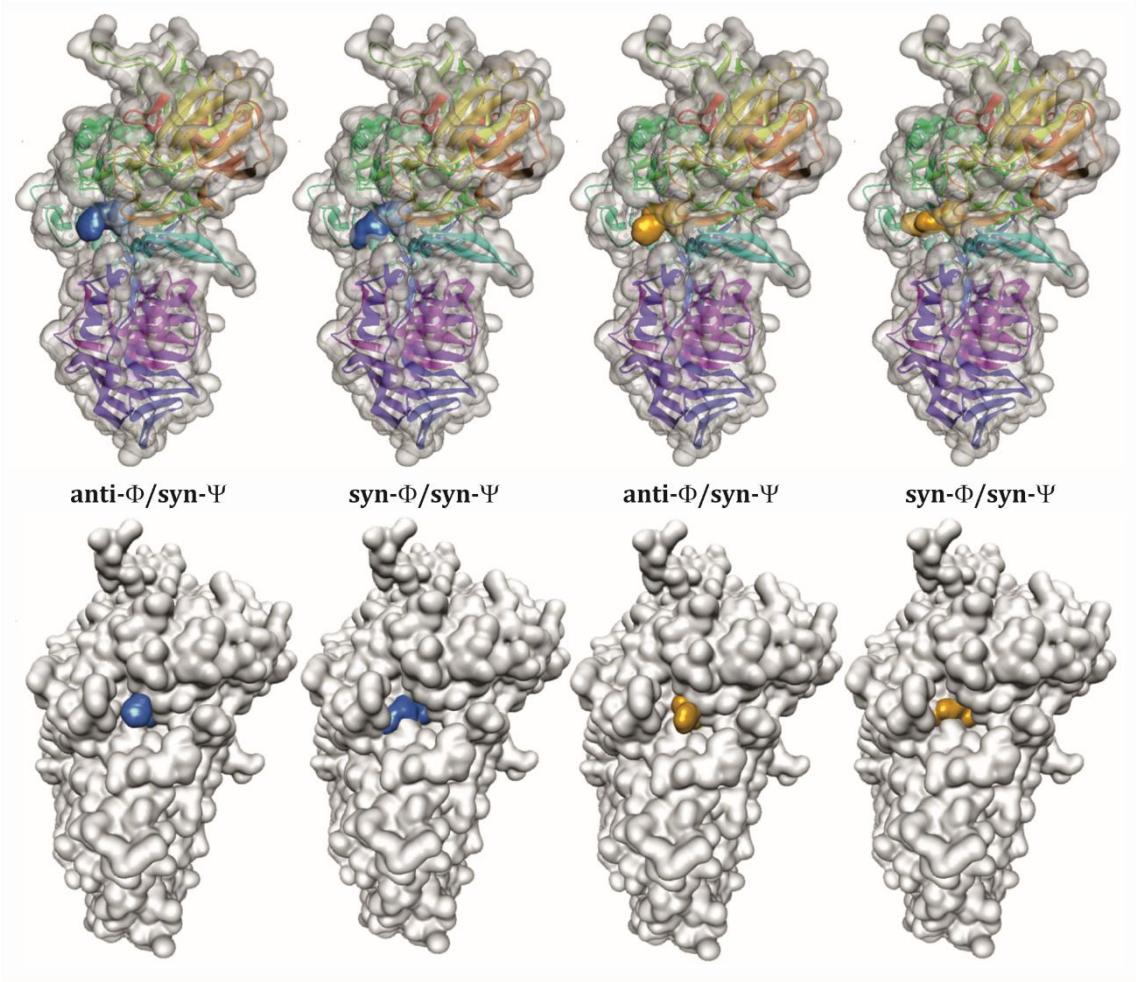


Fig 19. Surface representation of *E. coli* β-galactosidase in complex with the *anti*-Φ/*syn*-Ψ and *syn*-Φ/*syn*-Ψ conformers of **2*S*** (in blue) and **2*R*** (in yellow). The proper orientation of the aglycon acting as leaving group is observed only for the *anti*-Φ/*syn*-Ψ conformers.

3.3. CONCLUSIONS

Four sulfoxide glycomimetics have been studied. The conformational analysis of the acetylated and deacetylated derivatives has revealed a different conformational behavior depending on the stereochemistry of the sulfoxides. Furthermore, although both **2*R*** and **2*S*** derivatives act as moderate *E. coli* β-galactosidase inhibitors, the inhibition potency also depends on the stereochemistry at the sulfur atom. Despite the inhibition activity, both sulfoxides are hydrolyzed, albeit at very different rates. The high flexibility of **2*S*** plays an important role in the molecular recognition process and a conformational selection process takes place. The combination of NMR experimental data with modeling procedures explains the different rates of hydrolysis.

3.4. EXPERIMENTAL

NMR Spectroscopy: For **1*R*** and **1*S*** sulfoxides, NMR experiments were acquired at 298K using Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. The concentration employed was 5 mM for both compounds in $CDCl_3$. 1D ^{1}H NMR spectra, TOCSY (70 ms mixing time), NOESY (600 ms mixing time) and ^{1}H-^{13}C HSQC experiments were acquired, in order to assign all NMR signals. For 1D ^{1}H, TOCSY, NOESY and ^{1}H-^{13}C HSQC, the zgesp, dipsi2esgpph, noesyesgpph and hsqcedetgp sequences were employed. In the case of **2*R*** and **2*S*** sulfoxides, NMR experiments were acquired at 310K using Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. The concentration employed in the free-ligand experiments was 5 mM in sodium phosphate buffer (20 mM, pH 7.2) for both compounds. 1D ^{1}H NMR spectra, TOCSY (70 ms mixing time), NOESY (600 ms mixing time) and ^{1}H-^{13}C HSQC experiments were acquired, in order to assign all NMR signals. For 1D ^{1}H, TOCSY, NOESY and ^{1}H-^{13}C HSQC, the zgesp, dipsi2esgpph, noesyesgpph and hsqcedetgp sequences were employed. STD-NMR experiments (std2 sequence) were performed with a 300:1 ligand/enzyme molar ratio. An off-resonance frequency of 100 ppm and an on-resonance frequency of –1 ppm were applied, with 2 s of saturation time. Enzyme samples were recorded in sodium phosphate buffer (20 mM, pH 7.2) with 1 mM $MgCl_2$.

The NMR assignments of **1*R***, **1*S***, **2*R*** and **2*S*** are given below:

	1*R*		**1*S***	
Proton	**^{1}H Chemical shift (δ ppm)**	**^{13}C Chemical shift (δ ppm)**	**^{1}H Chemical shift (δ ppm)**	**^{13}C Chemical shift (δ ppm)**
H_1	4.88	97.63	4.81	97.35
H_2	4.05	64.47	4.00	64.68
H_{3A}	2.23	25.97	2.26	28.26
H_{3B}	2.49		2.16	
H_4	3.63	52.68	3.51	53.17
H_{5A}	4.14	58.60	4.13	57.57
H_{5B}	3.52		4.05	
$H_{1'}$	4.29	87.07	4.38	91.98
$H_{2'}$	5.72	63.86	5.59	65.88
$H_{3'}$	5.21	71.87	5.13	71.32
$H_{4'}$	5.49	66.78	5.49	66.84
$H_{5'}$	4.06	75.87	4.06	75.75
$H_{6'}$	4.18, 4.22	61.42	4.14, 4.18	61.23
OAc	2.03, 2.07, 2.12, 2.20	20,63, 20.71, 20.73, 20,77	2.02, 2.10, 2.19	20.59, 20.81, 20.70
CH_2 OBn	4.64, 4.82	70.11	4.63, 4.88	69.92
Bn	7.38	128.30	7.39	128.23

Table 3. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of **1*R*** and **1*S***.

	2*R*		**2*S***	
Proton	**^{1}H Chemical shift (δ ppm)**	**^{13}C Chemical shift (δ ppm)**	**^{1}H Chemical shift (δ ppm)**	**^{13}C Chemical shift (δ ppm)**
H_1	4.95	97.22	4.94	97.36
H_2	4.08	63.61	4.08	63.81
H_{3A}	2.25	24.34	2.31	26.39
H_{3B}			2.01	
H_4	3.89	51.36	3.93	52.28
H_{5A}	4.16	57.23	4.19	56.35
H_{5B}	3.60		3.97	
$H_{1'}$	4.33	87.90	4.60	91.79
$H_{2'}$	4.01	64.82	4.02	66.29
$H_{3'}$	3.79	73.77	3.79	73.98
$H_{4'}$	4.00	68.58	4.00	68.60
$H_{5'}$	3.85	79.95	3.84	80.36
$H_{6'}$	3.74, 3.81	60.85	3.75, 3.80	61.24

CH_2 OBn	4.65, 4.80	69.92	4.68, 4.82	69.99
Bn	7.45	128.48	7.46	128.61

Table 4. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of **2*R*** and **2*S***.

Compound 1*R*: ^{1}H NMR (600 MHz, Chloroform-*d*) δ 7.38 (dt, *J* = 14.2, 7.2 Hz, 5H), 5.72 (t, *J* = 10.0 Hz, 1H), 5.49 (d, *J* = 3.6 Hz, 1H), 5.21 (dd, *J* = 10.1, 3.4 Hz, 1H), 4.88 (d, *J* = 3.2 Hz, 1H), 4.82 (d, *J* = 11.8 Hz, 1H), 4.64 (d, *J* = 11.7 Hz, 1H), 4.29 (d, *J* = 10.0 Hz, 1H), 4.25 – 4.02 (m, 5H), 3.63 (p, *J* = 3.8 Hz, 1H), 3.54 – 3.48 (m, 1H), 2.49 (dt, *J* = 14.4, 4.6 Hz, 1H), 2.27 – 2.02 (m, 13H).

Compound 1*S*: ^{1}H NMR (600 MHz, Chloroform-*d*) δ 7 7.42 – 7.35 (m, 5H), 5.59 (t, *J* = 10.1 Hz, 1H), 5.49 (d, *J* = 3.3 Hz, 1H), 5.13 (dd, *J* = 9.9, 3.3 Hz, 1H), 4.88 (d, *J* = 11.7 Hz, 1H), 4.81 (d, *J* = 2.9 Hz, 1H), 4.63 (d, *J* = 11.8 Hz, 1H), 4.38 (d, *J* = 10.2 Hz, 1H), 4.21 – 4.10 (m, 3H), 4.09 – 4.04 (m, 2H), 4.00 (dp, *J* = 8.6, 4.9 Hz, 1H), 3.51 (p, *J* = 4.5 Hz, 1H), 2.30 – 1.99 (m, 14H).

Compound 2*R*: ^{1}H NMR (600 MHz, Deuterium Oxide) δ 7.48 – 7.42 (m, 5H), 4.95 (d, *J* = 3.4 Hz, 1H), 4.33 (d, *J* = 10.2 Hz, 1H), 4.16 (dd, *J* = 14.0, 3.2 Hz, 1H), 4.09 – 3.95 (m, 3H), 3.90 – 3.71 (m, 5H), 3.59 (dt, *J* = 14.3, 1.9 Hz, 1H), 2.31 – 2.20 (m, 2H).

Compound 2*S*: ^{1}H NMR (600 MHz, Deuterium Oxide) δ 7.50 – 7.43 (m, 5H), 4.94 (d, *J* = 3.4 Hz, 1H), 4.60 (s, 1H), 4.19 (dd, *J* = 13.0, 2.9 Hz, 1H), 4.10 – 3.99 (m, 3H), 3.97 (t, *J* = 2.1 Hz, 1H), 3.93 (dd, *J* = 7.5, 4.3 Hz, 1H), 3.87 – 3.72 (m, 4H), 2.31 (ddd, *J* = 15.4, 11.1, 4.7 Hz, 1H), 2.01 (dt, *J* = 15.4, 4.9 Hz, 1H).

Molecular Mechanics Calculations: A coordinate scan protocol was performed by using Macromodel,[14] implemented in the Maestro suite of programs. MM3* force field was used to obtain different conformations in order to select those who agree with the NMR spectroscopic data. The calculation method used the PRCG protocol by employing an energy-minimization process. Full geometry optimization for the MM3* structures was performed with the Gaussian 03 program,[15] using the HF/6-31+G(d) basis set.[16]

Molecular Dynamics Simulations: 3 ns MD simulations of the compounds in the free state were performed with Amber 12[17] (ff99SB and GAFF force fields), in explicit TIP3P water at 298 K.

Docking Procedure: 3D models of **1*S*** and **1*R*** in the binding site of the enzyme were built by using Macromodel.[14] The structure of *E. coli* β-galactosidase in complex with galactose (PDB 1JZ7) was used as starting model. The galactose residue of both molecules were superimposed, and the resulting structures were refined by energy minimization with the AMBER force field as implemented in Macromodel; several steps of the Polak–Ribière conjugate gradient (PRCG) were used until the energy gradient become smaller than 0.05 kJ/mol/Å.

3.5. REFERENCES

[1] H. Driguez, *ChemBioChem* **2001**, *2*, 311–318.
[2] A. J. Cagnoni, M. L. Uhrig, O. Varela, *Bioorg. Med. Chem.* **2009**, *17*, 6203–6212.
[3] V. E. Manzano, M. L. Uhrig, O. Varela, *Org. Biomol. Chem.* **2012**, *10*, 8884.
[4] E. Repetto, C. Marino, O. Varela, *Bioorg. Med. Chem.* **2013**, *21*, 3327–3333.
[5] L. Calle, V. Roldós, F. J. Cañada, M. L. Uhrig, A. J. Cagnoni, V. E. Manzano, O. Varela, J. Jiménez-Barbero, *Chem. - A Eur. J.* **2013**, *19*, 4262–4270.
[6] "Nexium Control | European Medicines Agency," can be found under https://www.ema.europa.eu/en/medicines/human/EPAR/nexium-control
[7] N. Khiar, I. Alonso, N. Rodriguez, A. Fernandez-Mayoralas, J. Jimenez-Barbero, O. Nieto, F. Cano, C. Foces-Foces, M. Martin-Lomas, *Tetrahedron Lett.* **1997**, *38*, 8267–8270.
[8] J. P. Colomer, M. Á. Canales Mayordomo, B. Fernández de Toro, J. Jiménez-Barbero, O. Varela, *Eur. J. Org. Chem.* **2015**, *2015*, 1448–1455.
[9] J. P. Colomer, B. Fernández de Toro, F. J. Cañada, F. Corzana, J. Jiménez Barbero, Á. Canales, O. Varela, *Eur. J. Org. Chem.* **2016**, *2016*, 5117–5122.
[10] Mestrelab Research S.L., **2012**, version 8.1.0-11315.
[11] A. García-Herrero, E. Montero, J. L. Muñoz, J. F. Espinosa, A. Vián, J. L. García, J. L. Asensio, F. J. Cañada, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **2002**, *124*, 4804-4810.
[12] B. W. Matthews, *C. R. Biol.* **2005**, *328*, 549–556.
[13] D. H. Juers, T. D. Heightman, A. Vasella, J. D. McCarter, L. Mackenzie, S. G. Withers, B. W. Matthews, *Biochemistry* **2001**, *40*, 14781-14794.
[14] Schrödinger, **2013**.
[15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, Gaussian 03, revision E.01, Gaussian, Inc., Wallingford, CT, **2004**.
[16] G. I. Csonka, *J. Mol. Struct. THEOCHEM* **2002**, *584*, 1–4.

[17] T. A. D. D. A. Case, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, A. W. G. J. Swails, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, S. R. B. X. Wu, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 12, University of California, San Francisco, **2012**.

CHAPTER 4

N-GLYCANS RECEPTORS OF THE INFLUENZA A VIRUS: NOVEL METHODOLOGIES FOR THEIR CONFORMATIONAL ANALYSIS AND CHARACTERIZATION OF THEIR INTERACTIONS WITH VIRAL HEMAGGLUTININS.

4.1. INTRODUCTION

Among membrane-enveloped viruses, influenza virus belongs to the Orthomyxoviridae family. There are four genera of influenza viruses (A, B, C and D) of which three infect humans (A, B and C). However, while B and C viruses are only found in humans, influenza A virus (IAV) can also infect other animals such as birds.[1,2]

Seasonal flu is the respiratory disease provoked by these viruses and is one of the major threats to human health.[3,4]

Influenza viruses are transmitted among individuals within airborne respiratory droplets. The viral infection begins with the attachment of the virus to cell-surface receptors. In the case of IAV, two glycoproteins located on the virus membrane, Hemagglutinin (HA) and Neuraminidase (NA) recognize sialic acid.[5] HA is responsible for initiating the infection by binding to sialylated glycans located on the surface of the cells of the respiratory tract.[2,6]

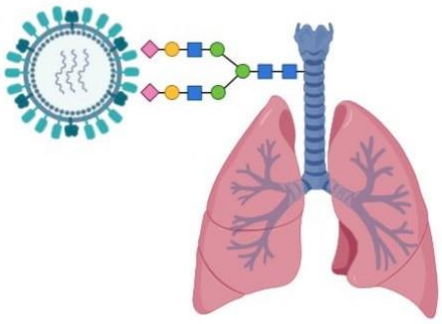


Fig 1. Schematic representation of the Influenza virus hemagglutinin interaction with the sialylated N-glycans. Created with www.BioRender.com

After the HA-mediated interaction, the influenza life cycle continues by releasing the genetic material inside the host cell, replication, and assembly of the new virions ready for viral release. When budding is completed, the HA remains attached to the sialylated N-glycans of the host cells until the NA cleaves the sialylα2-3- or sialylα2-6-galactose linkage (see below). The sialidase activity also removes the sialic acid residues of the new virions to prevent aggregation.

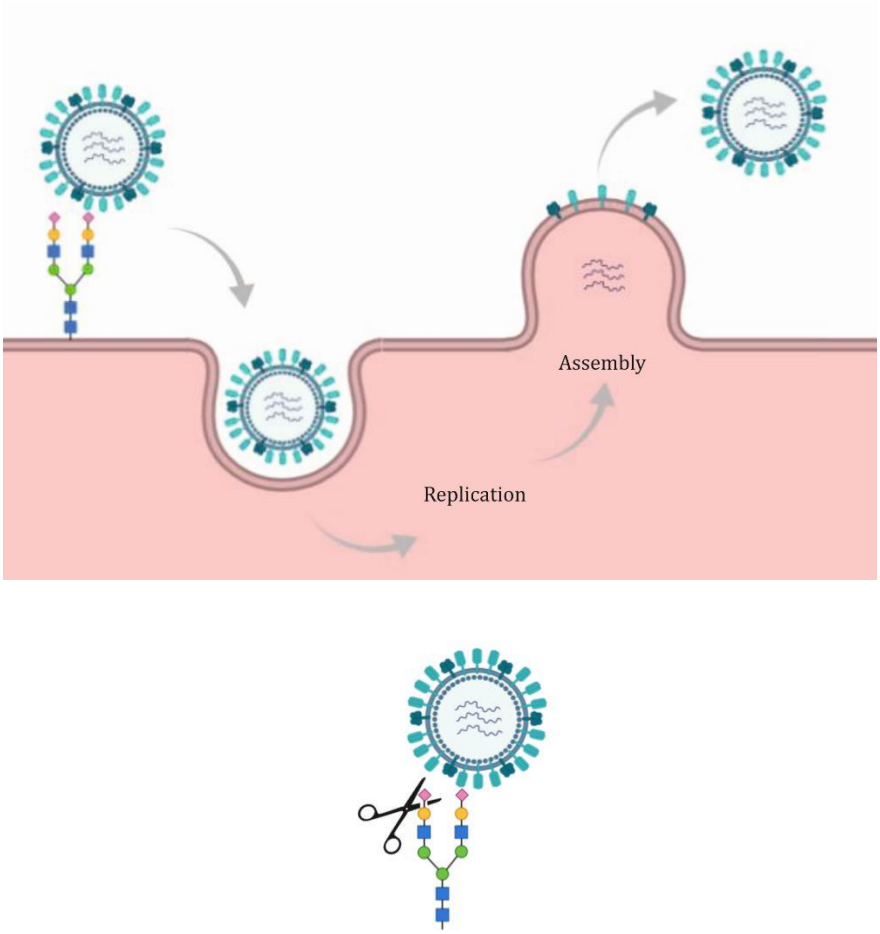


Fig 2. Top: Schematic representation of the Influenza virus cycle. **Bottom:** Schematic representation of the cleavage of the sialylα2-3- or sialylα2-6-galactose linkage by the NA. Created with www.BioRender.com

In absence of neuraminidase activity, the new virions remain together and do not spread and continue infecting. Given this observation, and following a structure-based drug design approach, two drugs were designed to block the NA and therefore act as anti-influenza drugs (Zanamivir and Oseltamivir).[2,7-9]

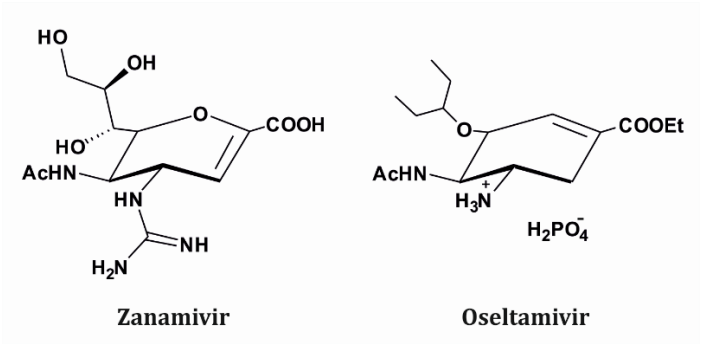


Fig 3. Structure of Zanamivir (left) and Oseltamivir (right).

Influenza HA is a type I integral membrane glycoprotein, where the glycosylation occurs by post-translational modifications. A single precursor is cleaved by cellular proteases and renders two polypeptides, the globular domain (HA1), which interacts with the sialic acid residue, and a stalk-like domain (HA2), which facilitates the membrane fusion. Both subunits form a disulfide bond-linked complex. The receptor binding site (RBS) is located at the HA1 domain and is conserved in all influenza subtypes. The HA forms trimers of three identic monomers, each of which presents one RBS able to interact with the sialylated N-glycans receptors.[6,7,10-12]

The HA receptor specificity and the consequent species infectivity were studied by using hemagglutination studies of avian and human H3N2 subtypes in the presence of cells with α2-3 and α2-6 sialic acid-galactose linkage. The results showed the preference for the α2-3 linkage in the case of avian viruses and the α2-6 linkage in human viruses.[7,10,13] This minor difference in the recognition of the glycan is a key factor in the species barrier, since α2-6-linked sialic acids are the most abundant in human epithelial cells, while the α2-3 analogues are those in avian epithelial cells, thus preventing the infection of humans by avian viruses. Another barrier is constituted by the mucins located on the upper respiratory tract epithelium. They are enriched with α2-3 glycans, which retain avian influenza viruses and prevent their attachment to epithelial cells.[14]

Influenza A viruses are classified according to the antigenic proteins of the two major glycoproteins, HA (H1-18) and NA (N1-11), and they generally circulate in avian species although occasionally infect and adapt to humans, causing worldwide pandemics. As previously said, one barrier to transmission of avian viruses in humans is receptor specificity.[10] However, at a certain moment, some avian-origin influenza viruses have crossed this barrier and infected humans. Four pandemics took place in the last decades:

H1N1	**1918** Spanish pandemic (50 million deaths)[15] **2009** H1N1 pandemic (<0.5 million deaths)[16]
H2N2	**1957** Asian pandemic (1.1 million deaths)[17]
H3N2	**1968** Hong Kong pandemic (1 million deaths)[18]

Table 1. Pandemics recorded in the last decades.

A zoonotic virus has to overcome three barriers to infect humans: animal to human transmission, virus-cell interaction and human to human transmission.[19] Currently, H3N2 (1968) and H1N1 (2009) subtypes circulate within humans.

In the following figure, the reassortment and adaptations of pandemic influenza A viruses responsible of the four pandemics are represented.[20]

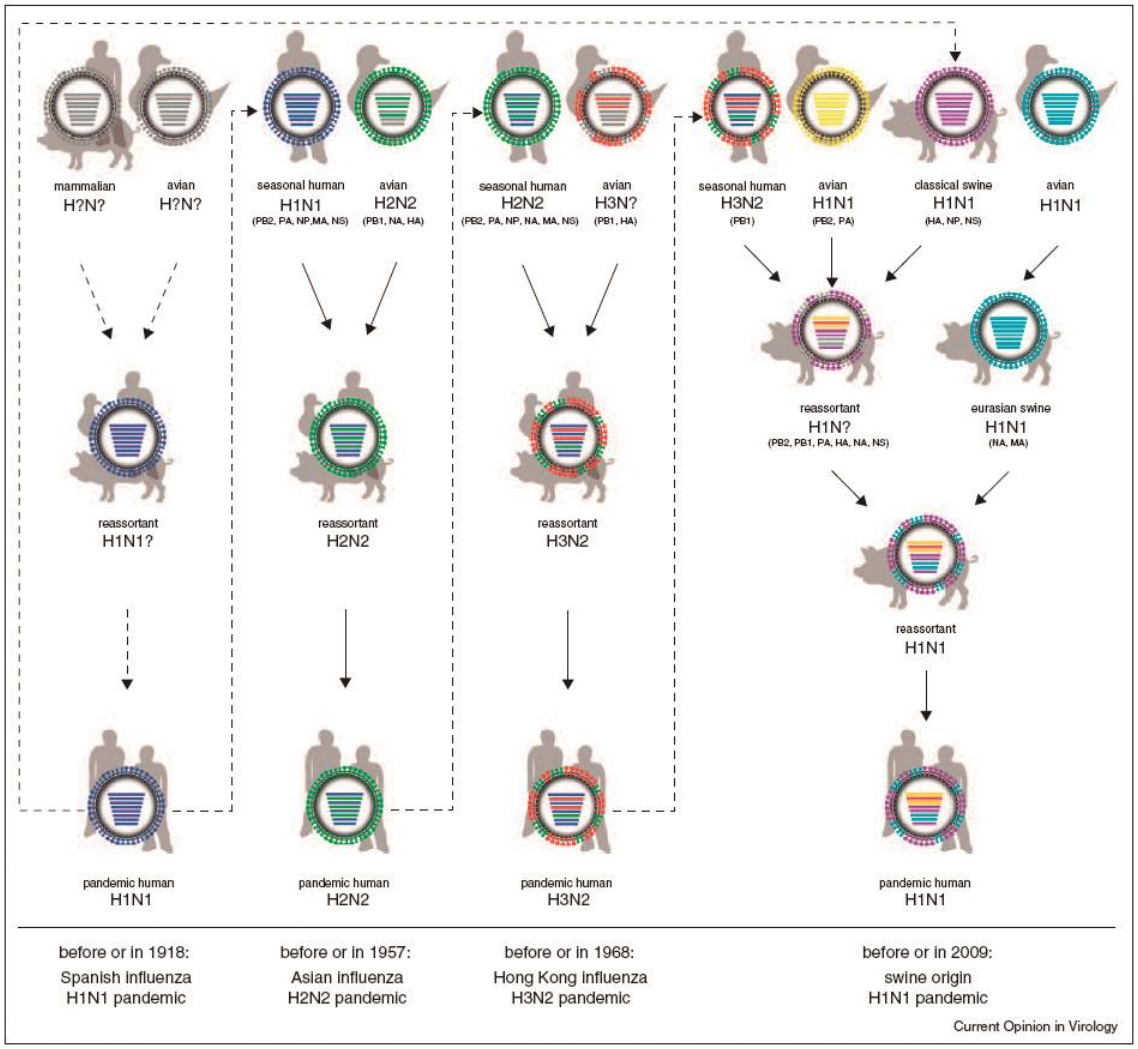


Fig 4. Scheme representation of the reassortment and adaptations of pandemic influenza A viruses. (Reprinted from Current opinion in virology, vol 1, Sorrel E.; Schrauwen, E.; Linster, M.; De Graaf, M.; Herfst, S.; Fouchier, R., 635-642, Predicting "airborne" Influenza viruses: (Trans-) mission impossible?, Copyright (2011), with permission from Elsevier).

The 1918 Spanish flu pandemic is believed to have been originated from spontaneous mutations in an avian influenza virus that enabled human to human transmission. However, the H1N1 pandemic of 2009 was mild compared to the 1918 pandemic, but what really worried the WHO was the rapid spread it presented.[7,21,22]

This fact highlights the importance of predicting the virulence of the new influenza strains and be able to develop a vaccine. In this context, is of crucial importance the understanding of the specificity basis.

As reported before, influenza HA is responsible for initiating the infection by binding to sialylated glycans located on the surface of the cells of the respiratory tract. The mutation of only two amino acids permits the avian viruses to switch the specificity from α2-3 linkage to α2-6, permitting the infection of humans and therefore initiate a pandemic infection.[23-25]

In humans, the glycan profiling of the epithelial cells from the upper respiratory tract shows that, although a substantial diversity of structures is found, there is a predominant expression of α2-6 sialylated branched N-glycans with polyLacNAc units (Galβ1-4GlcNAcβ1-3 tandem repeats).[26] In fact, recent studies. By using glycoarray methodology, have pointed out that not only the sialic acid is important in the recognition, but also the length of the glycan chain is a key factor. In this context, certain H3N2 variants have evolved specificity to recognize long chain glycans with polyLacNAc units.[27] Consequently, not only the α2-6 linkage is important but the glycan complexity and structural and chemical aspects are also key factors in the influenza A virus infection.

In this context, the work described in this chapter has been focused on applying novel NMR methodologies to obtain detailed information on the key glycan binding epitope for HA recognition, with the final aim of better understand the molecular basis of influenza virus infection.

4.2. CONFORMATIONAL ANALYSIS OF N-GLYCANS RECEPTORS OF THE INFLUENZA A VIRUS

4.2.1. Introduction

Long-chain biantennary N-glycans, as is the case of the influenza A virus receptors, are extremely complex molecules. However, the advances in synthetic methodologies are now permitting to obtain such complex structures in sufficient amounts to perform detailed structural and molecular recognition studies. However, the intrinsic flexibility and the presence of many similar monosaccharides and repeating unit within the same chemical structure, poses a tremendous challenge to unravel the conformational, dynamic and recognition features of these molecules.

Although NMR spectroscopy is successfully used for studying carbohydrates, the study of long-chain multiantenna N-glycans using the standard NMR experiments is basically impossible. The chemical equivalence of many of the NMR-active nuclei in these molecules precludes gaining unambiguous information on branch specificity, molecular recognition features and epitope characterization. This is particularly overstated for the influenza A virus receptors, which present pseudo-symmetry and multiple LacNAc repeats at each branch (figures 5A, 5B).

In this context, the use of paramagnetic NMR and specifically, the employment of N-glycans bearing lanthanide-binding tag (LBT), has emerged in the last years as a powerful tool that allows the NMR-based study of the conformational behavior of complex glycans.[28,29]

4.2.2. Results and discussion

As mentioned above, the typical glycan structures that are recognized by HA are gathered in figure 5A. The work presented herein has been carried out with two N-glycans (**1**, **2**) synthesized by the group of Prof. James Paulson at The Scripps Research Institute in San Diego, California.[30]

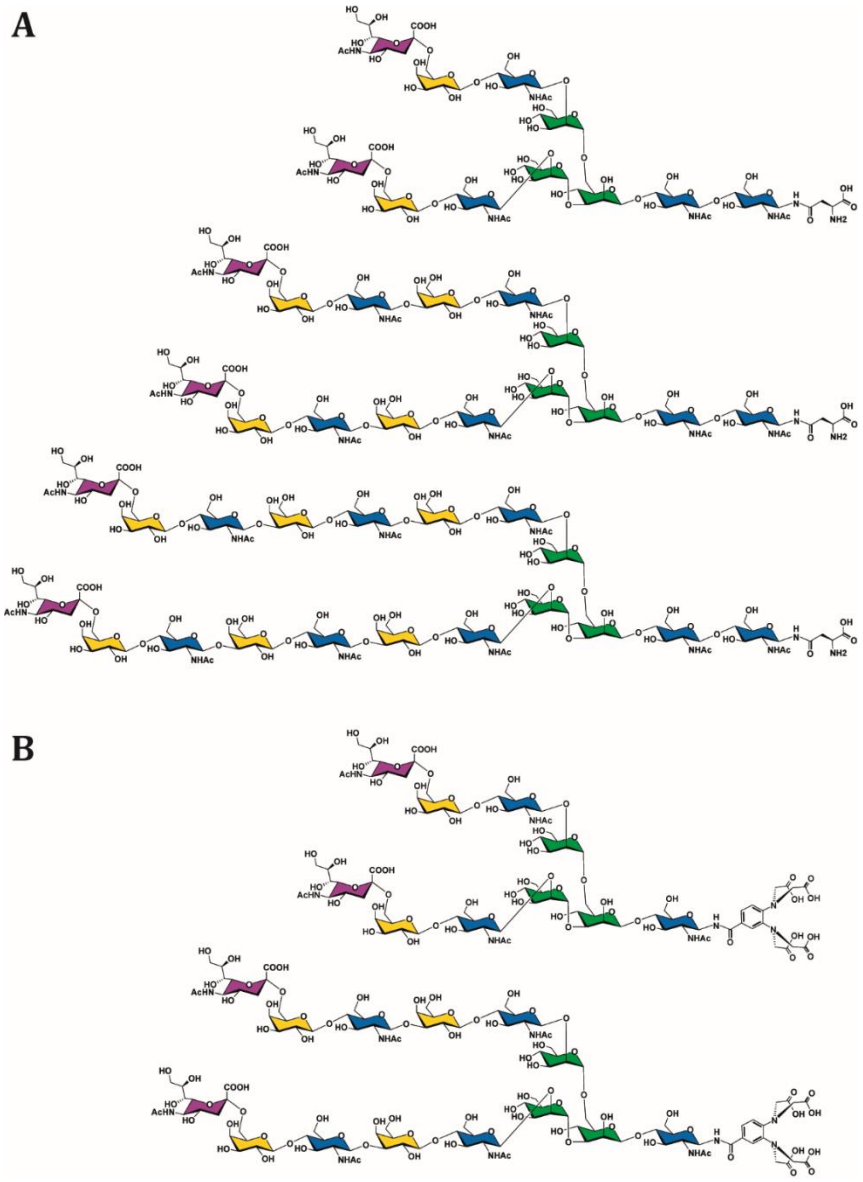


Fig 5. A) Schematic structure of the typical N-glycans that are recognized by HA. B) Schematic structure of the N-glycans **1** (top) and **2** (bottom). Saccharide units are colored according to the Symbol Nomenclature for Glycans (SNFG).[31]

1 and **2** (figure 5B) are α2-6 sialylated biantennary N-glycans containing one or two LacNAc repeats, respectively. In order to characterize and accomplish the conformational study of both N-glycans, they were conjugated with an LBT. After initial trials with the regular glycans containing the chitobiose unit attached to the Asn moiety, decorated with a LBT, we decided to remove the GlcNAc-Asn fragment to be able to detect long-range significant perturbations (pseudo contact shifts) at the non-reducing end provoked by the presence of the paramagnetic ion at the LBT.

For the NMR studies, two ^{1}H-^{13}C HSQC NMR spectra of the decasaccharide N-glycan **1** were acquired (figure 6). In the reference experiment, the LBT was loaded with lanthanum (diamagnetic conditions), while dysprosium (paramagnetic conditions) was employed to generate the PCS. Fittingly, the presence of Dy^{3+} at the LBT permitted differentiation of the signals of both branches of the N-glycan. In fact, the four sugars at each branch were distinguished. Strikingly, the comparison of the diamagnetic and paramagnetic ^{1}H-^{13}C HSQC NMR spectra shows that the unique set of signals for every pair of Neu5Ac, Gal and GlcNAc units in the La^{3+}-containing sample is split into two distinct sets of cross peaks in the presence of Dy^{3+}.

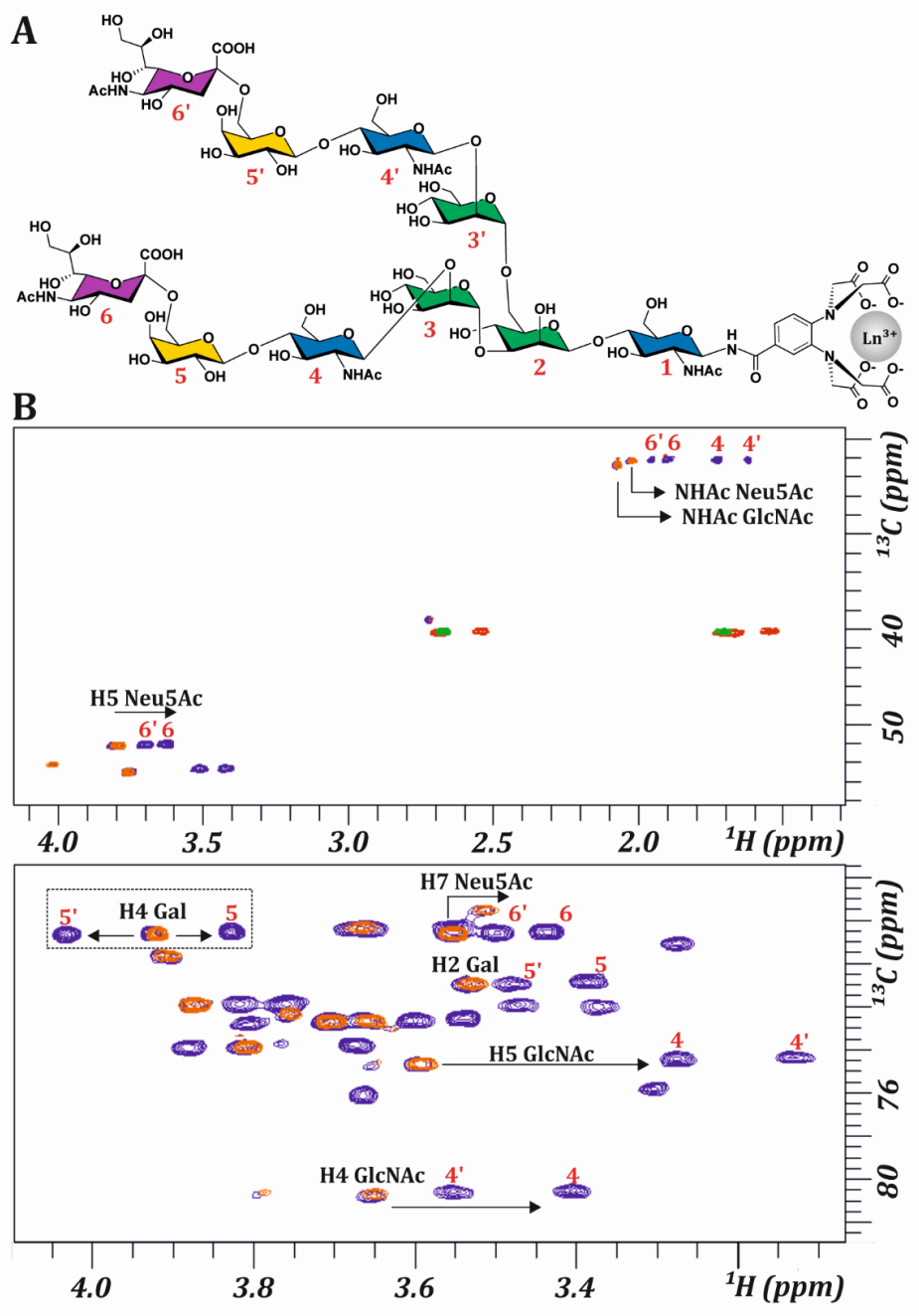


Fig 6. A) Schematic view of **1** along with the numbering for each monosaccharide unit. B) Superimpositions of selected sections of the edited ^{1}H-^{13}C HSQC spectra of **1** loaded with lanthanum (diamagnetic, CH and CH_3 in orange, CH_2 in green) and with dysprosium (paramagnetic, CH and CH_3 in blue, CH_2 in red). The split signals are highlighted: the differentiation of both branches is achieved.

The same protocol was applied for N-glycan **2**. In this case, the challenge is even higher, given its tetradecasaccharide structure, which presents two LacNAc repetitions at each arm. The analysis of the corresponding ^{1}H-^{13}C HSQC NMR spectra of **2** complexed with Dy^{3+} also permitted the identification of different NMR signals for every sugar unit.

Remarkably, even the Neu5Ac external residues, located far away from the paramagnetic lanthanide, can be distinguished. For instance, H7 of one Neu5Ac unit was estimated to be located at ca. 37 Å from the Dy^{3+} ion. Moreover, the signals for all Gal and GlcNAc units could be distinguished (figure 7).

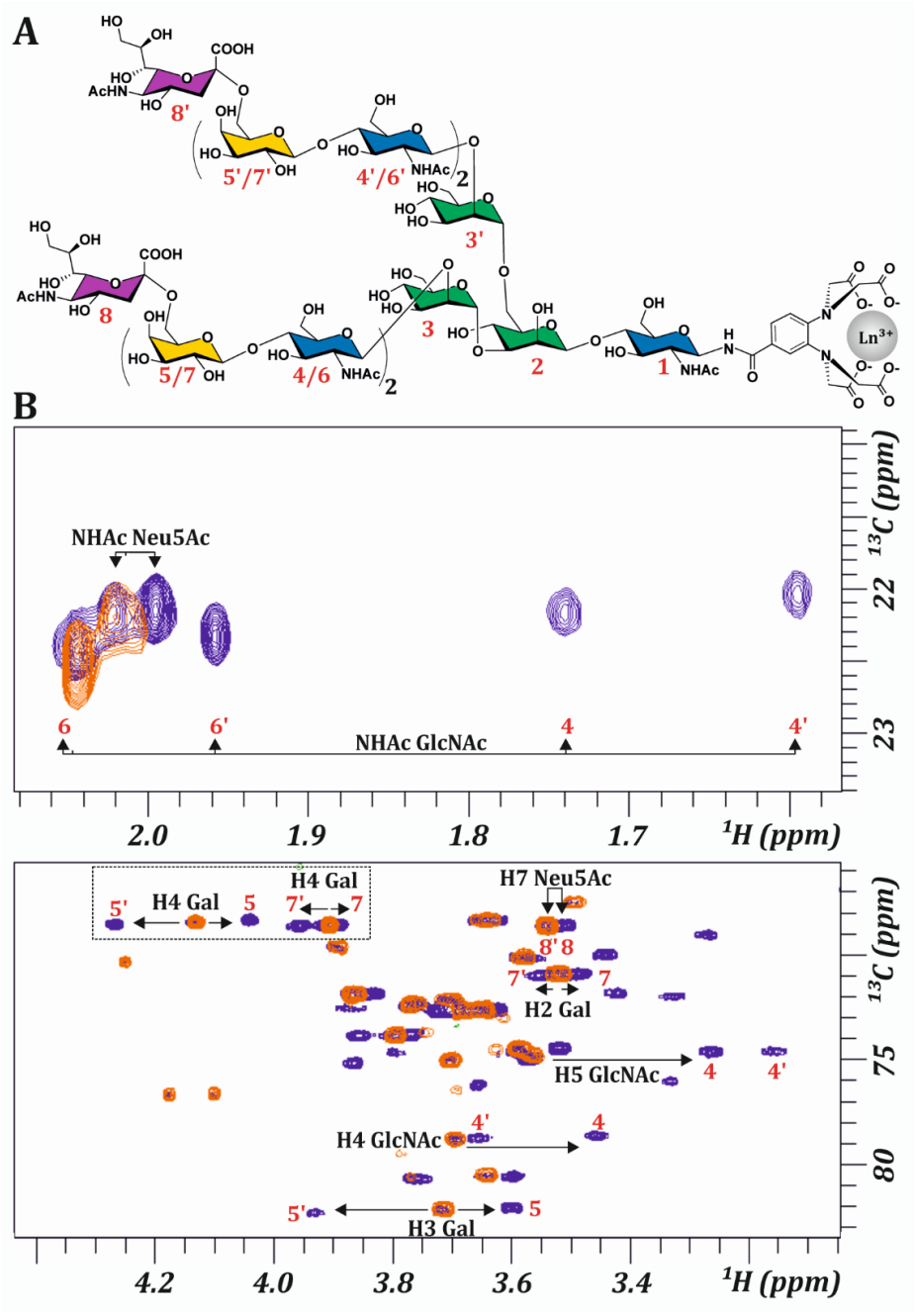


Fig 7. A) Schematic view of N-glycan **2** along with the numbering for each monosaccharide unit. B) Superimpositions of selected sections of ^{1}H-^{13}C HSQC spectra of N-glycan **2** loaded with lanthanum (diamagnetic, in orange) and with dysprosium (paramagnetic in blue). Again, the split signals are highlighted, permitting the unambiguous distinction of both branches.

The experimental Pseudo Contact Shift values (PCS) were obtained as the difference in chemical shifts measured between the ^{1}H-^{13}C HSQC spectra acquired in the presence of La^{3+} (diamagnetic conditions) and Dy^{3+} (paramagnetic conditions). As stated above, PCS depend on the distance and orientation of the observed nucleus and the paramagnetic ion.

Interestingly, for N-glycans **1** and **2**, different PCSs behavior was observed for some proton signals. For instance, for **2**, H4 of residues Gal 5' and Gal 7' displayed downfield shifting, while H4 of Gal 5 and Gal 7 shifted upfield (figure 7, panel B, lower ^{1}H-^{13}C HSQC spectra).

The conformational behavior of both **1** and **2** N-glycans was deduced by comparison of the experimental PCS values with those estimated by back-calculating the expected PCS values for different geometries of the N-glycans obtained by molecular modelling. These geometries mainly differed in the orientations around the Ψ/ω glycosidic torsions of the Manα (1-6) Manβ linkage, resulting in three conformers: extended conformers (Ψ 180°) with gg and gt rotamers (ω 60° and 180°, respectively), and the folded gg conformer (Ψ 90°, ω 60°) (figure 8).

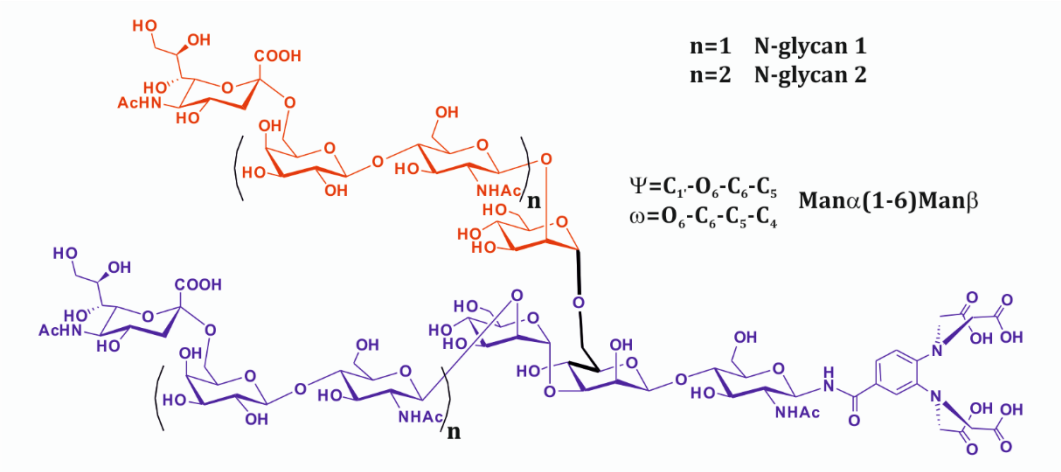


Fig 8. Structures of N-glycans **1** (n=1) and **2** (n=2). The definition of the torsion angles for the Manα (1-6) Manβ linkage is given.

No satisfactory fits were found between the experimental and the expected PCS values for any single conformation (Tables 2 and 3), suggesting the presence of a conformational distribution. For N-glycan **1**, the best fit was obtained for a two-state equilibrium composed by the extended conformers (Ψ 180°) with a major contribution (70%) of the gg rotamer (ω 60°) and a minor but significant presence (30%) of the gt one (ω 180°) (figure 9). The associated quality factor was excellent (0.16) (figure 11).

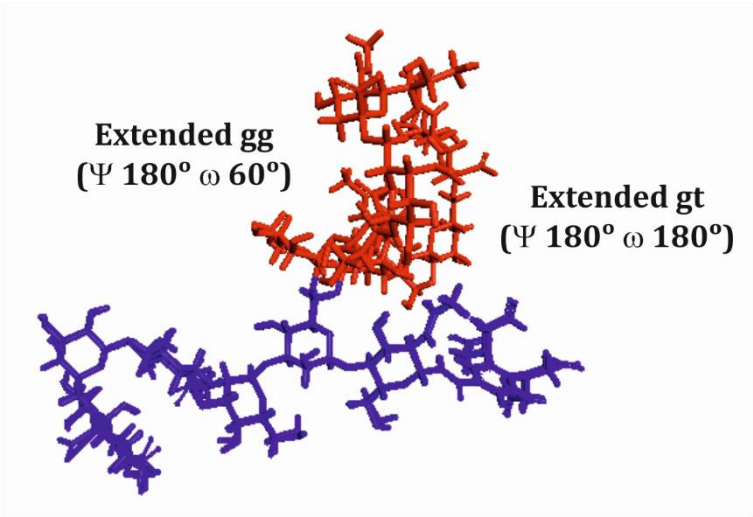


Fig 9. Superimposition of the extended gg and extended gt geometries of the Manα (1-6) Manβ linkage found in solution for N-glycan **1**, which provide the best fit to explain the experimental PCS values. The arm with the Manα (1-6) Manβ linkage is represented in red.

In contrast, the best fit for N-glycan **2** strongly suggests the existence of a different conformational distribution, with three participating conformers equally populated: extended gg (Ψ 180°, ω 60°), extended gt (Ψ 180°, ω 180°) and folded gg (Ψ 90°, ω 60°) (figure 10). The corresponding quality factor was also very good, 0.24 (figure 11).

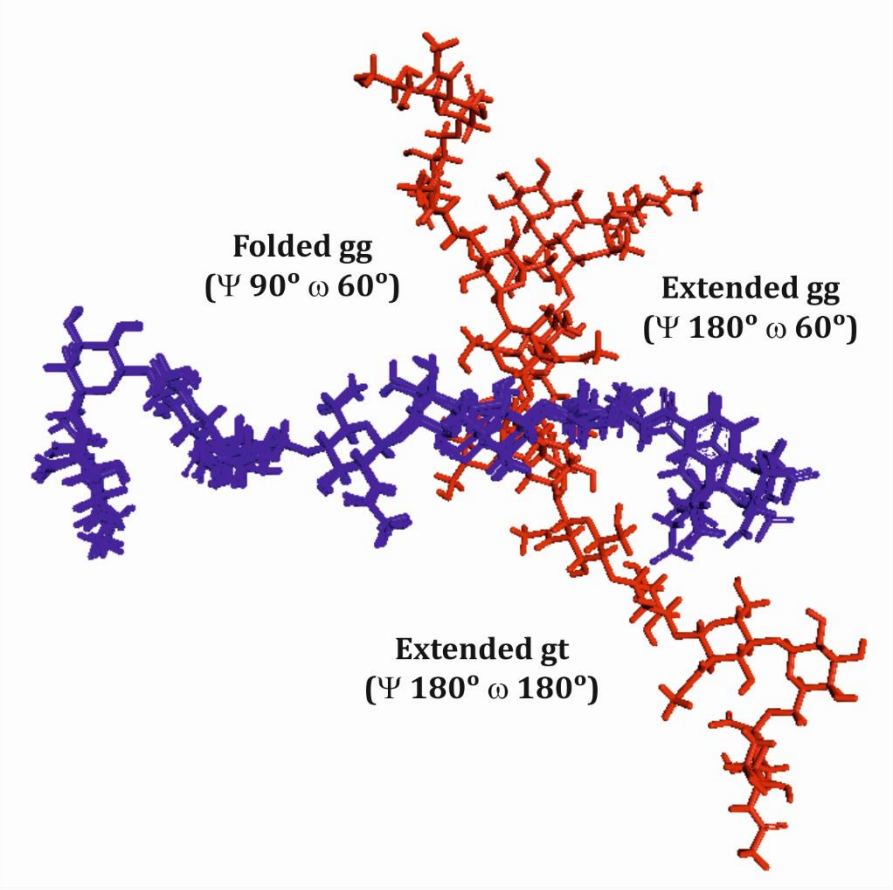


Fig 10. Superimposition of the extended gg, extended gt and folded gg geometries of the Manα (1-6) Manβ linkage found in solution for N-glycan **2**. The 1:1:1 conformation provided the best fit between the expected and the experimental PCS values. The arm with the Manα (1-6) Manβ linkage is represented in red.

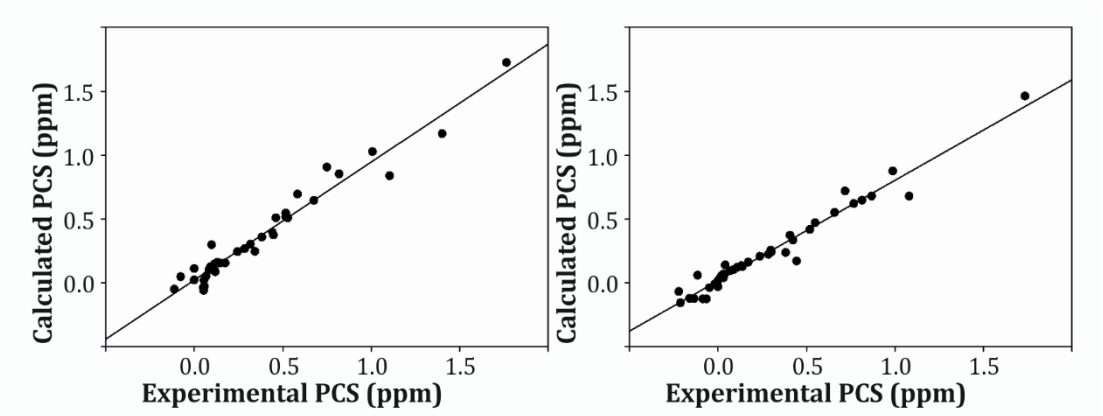


Fig 11. Correlation between the experimental PCS and the calculated values for a combination of the two conformations (extended gg:extended gt 70:30) for the N-glycan **1** (left panel) and for the three equally populated conformations of N-glycan **2** (right panel).

	Experimental	Calculated				
		Individual fittings				Combined
		gg (Ψ 180) Q factor 0.17	gt (Ψ 180) Q factor 0.26	gg (Ψ 90) Q factor 0.34	gg (Ψ 60) Q factor 0.26	Q factor 0.16
H_5 Neu5Ac 6'	0.08	0.11	0.09	-0.11	-0.01	0.10
H_6 Neu5Ac 6'	0.00	0.03	0.00	-0.14	-0.09	0.02
H_7 Neu5Ac 6'	0.06	0.02	0.02	-0.03	0.05	0.02
H_8 Neu5Ac 6'	0.06	-0.05	0.03	0.07	0.07	-0.03
NHAc Neu5Ac 6'	0.07	0.07	0.02	-0.46	-0.11	0.05
H_2 Gal 5'	0.05	-0.08	-0.01	0.17	0.13	-0.06
H_3 Gal 5'	0.05	0.00	-0.12	0.08	0.04	-0.04
H_4 Gal 5'	-0.11	-0.04	-0.08	0.05	0.01	-0.05
H_5 Gal 5'	-0.08	0.05	0.05	0.04	-0.01	0.05
H_1 GlcNAc 4'	0.58	0.71	0.67	0.61	0.52	0.69
H_3 GlcNAc 4'	0.38	0.35	0.37	0.44	0.37	0.36
H_4 GlcNAc 4'	0.10	0.22	0.49	0.20	0.15	0.30
H_5 GlcNAc 4'	0.46	0.43	0.69	0.37	0.31	0.51
NHAc GlcNAc 4'	0.45	0.42	0.26	0.83	0.83	0.37
H_1 Man 3'	0.75	0.98	0.73	1.07	1.01	0.91
H_2 Man 3'	0.82	0.73	1.14	0.57	0.55	0.85
H_4 Man 3'	0.52	0.42	0.83	0.33	0.25	0.55
H_5 Neu5Ac 6	0.18	0.16	0.15	0.16	0.14	0.16
H_6 Neu5Ac 6	0.00	0.11	0.11	0.12	0.10	0.11
H_7 Neu5Ac 6	0.12	0.09	0.09	0.12	0.10	0.09
H_8 Neu5Ac 6	0.11	0.10	0.11	0.14	0.12	0.10
NHAc Neu5Ac 6	0.12	0.09	0.07	0.10	0.05	0.09
H_2 Gal 5	0.14	0.15	0.17	0.17	0.18	0.15
H_3 Gal 5	0.12	0.14	0.16	0.14	0.15	0.15
H_4 Gal 5	0.09	0.12	0.13	0.13	0.13	0.12
H_5 Gal 5	0.13	0.16	0.17	0.15	0.15	0.16
H_1 GlcNAc 4	0.44	0.40	0.38	0.40	0.40	0.39
H_3 GlcNAc 4	0.28	0.27	0.27	0.31	0.30	0.27
H_4 GlcNAc 4	0.24	0.24	0.25	0.23	0.24	0.24
H_5 GlcNAc 4	0.32	0.30	0.31	0.29	0.31	0.30
NHAc GlcNAc 4	0.34	0.24	0.25	0.34	0.32	0.25
H_1 Man 3	0.68	0.66	0.60	0.59	0.65	0.65
H_2 Man 3	0.53	0.51	0.51	0.43	0.48	0.51
H_4 Man 3	0.52	0.52	0.51	0.38	0.43	0.52
H_1 Man 2	1.76	1.81	1.52	1.41	1.61	1.72
H_2 Man 2	1.40	1.16	1.19	0.86	0.99	1.17
H_3 Man 2	1.01	1.07	0.91	0.90	1.01	1.03
H_4 Man 2	1.10	0.85	0.80	0.78	0.85	0.84

Table 2. Experimental and calculated PCSs obtained from the different conformations of N-glycan **1**.

	Experimental	Calculated				
		Individual fittings				Combined
		gg (Ψ 180) Q factor 0.32	gt (Ψ 180) Q factor 0.34	gg (Ψ 90) Q factor 0.25	gg (Ψ 60) Q factor 0.24	Q factor 0.22
H_5 Neu5Ac 8'	0.00	-0.09	0.00	-0.01	0	-0.03
H_7 Neu5Ac 8'	0.00	-0.07	-0.02	0.00	0	-0.03
H_1 Gal 7'	-0.09	-0.04	-0.26	-0.09	-0.9	-0.13
H_2 Gal 7'	-0.02	-0.04	0.05	-0.05	-0.05	-0.01
H_4 Gal 7'	-0.05	-0.04	0.00	-0.08	-0.07	-0.04
H_5 Gal 7'	-0.07	-0.04	-0.26	-0.09	-0.09	-0.13
H_5 GlcNAc 6'	-0.21	-0.04	-0.38	-0.05	-0.07	-0.16
H_1 Gal 5'	-0.12	0.01	0.20	-0.03	-0.04	0.06
H_2 Gal 5'	0.00	-0.12	0.12	0.04	0.03	0.01
H_3 Gal 5'	-0.22	-0.04	-0.11	-0.05	-0.05	-0.07
H_4 Gal 5'	-0.14	-0.05	-0.19	-0.14	-0.12	-0.12
H_5 Gal 5'	-0.16	-0.01	-0.18	-0.18	-0.17	-0.12
H_1 GlcNAc 4'	0.55	0.34	0.61	0.45	0.36	0.44
H_3 GlcNAc 4'	0.38	0.07	0.36	0.28	0.21	0.28
H_4 GlcNAc 4'	0.04	0.07	0.37	-0.03	-0.04	0.21
H_5 GlcNAc 4'	0.41	0.15	0.71	0.25	0.20	0.13
NHAc GlcNAc 4'	0.44	0.13	0.23	0.15	0.49	0.35
H_1 Man 3'	0.72	0.67	0.64	0.85	0.80	0.70
H_2 Man 3'	0.77	0.46	1.00	0.40	0.41	0.62
H_5 Man 3'	0.81	0.39	0.93	0.62	0.86	0.73
H_5 Neu5Ac 8	0.03	0.04	0.06	0.04	0.04	0.05
H_7 Neu5Ac 8	0.03	0.03	0.05	0.03	0.03	0.04
H_1 Gal 7	0.03	0.06	0.07	0.07	0.06	0.06
H_2 Gal 7	0.04	0.05	0.07	0.06	0.06	0.06
H_4 Gal 7	0.02	0.04	0.05	0.05	0.05	0.05
H_5 Gal 7	0.02	0.05	0.06	0.05	0.05	0.05
H_5 GlcNAc 6	0.07	0.09	0.10	0.09	0.09	0.09
H_1 Gal 5	0.17	0.16	0.16	0.16	0.16	0.16
H_2 Gal 5	0.14	0.12	0.14	0.12	0.11	0.12
H_3 Gal 5	0.11	0.12	0.13	0.12	0.12	0.12
H_4 Gal 5	0.09	0.11	0.10	0.10	0.10	0.10
H_5 Gal 5	0.13	0.14	0.13	0.14	0.14	0.13
H_1 GlcNAc 4	0.42	0.35	0.32	0.33	0.34	0.34
H_3 GlcNAc 4	0.29	0.24	0.22	0.21	0.20	0.24
H_4 GlcNAc 4	0.24	0.21	0.20	0.21	0.21	0.22
H_5 GlcNAc 4	0.30	0.24	0.25	0.24	0.24	0.21
NHAc GlcNAc 4	0.30	0.35	0.21	0.20	0.16	0.24
H_1 Man 3	0.66	0.61	0.47	0.57	0.59	0.56
H_2 Man 3	0.52	0.42	0.39	0.44	0.45	0.42
H_5 Man 3	0.87	0.65	0.59	0.80	0.83	0.69
H_1 Man 2	1.73	1.53	1.22	1.63	1.78	1.51
H_3 Man 2	0.99	0.97	0.73	0.92	0.95	0.88
H_4 Man 2	1.08	0.67	0.67	0.69	0.72	0.69

Table 3. Experimental and calculated PCSs obtained from the different conformations of N-glycan **2**.

According to these results, we can deduce that there is a certain conformational restriction around Ψ dihedral angle for the N-glycan **1**, while **2**, the longest structure presenting 2 LacNAc repeats, is rather more flexible. The hairpin shape, formed by the Neu5Acα (2-6) Gal linkage[12,26] and confirmed by NOE signals between the H_{3ax} of the Neu5Ac unit and the acetyl group of the GlcNAc that belongs to the LacNAc unit attached to the Neu5Ac, somehow restricts the conformational flexibility, since the folded conformers may display steric clashes (figure 12).

Interestingly, the experimental data show that N-glycan **1** is less flexible than the analogous asialo biantennary N-glycan previously described using this methodology.[28] Thus, the presence of just an additional sialic acid unit at the end of the glycan chains modulate the conformational behavior of the molecule.

In addition, longer chains can explore a wider conformational space, resulting in a higher flexibility of the N-glycan. This flexibility can be a key factor to explain the preferential recognition of longer glycans by HA viral proteins.[27,32]

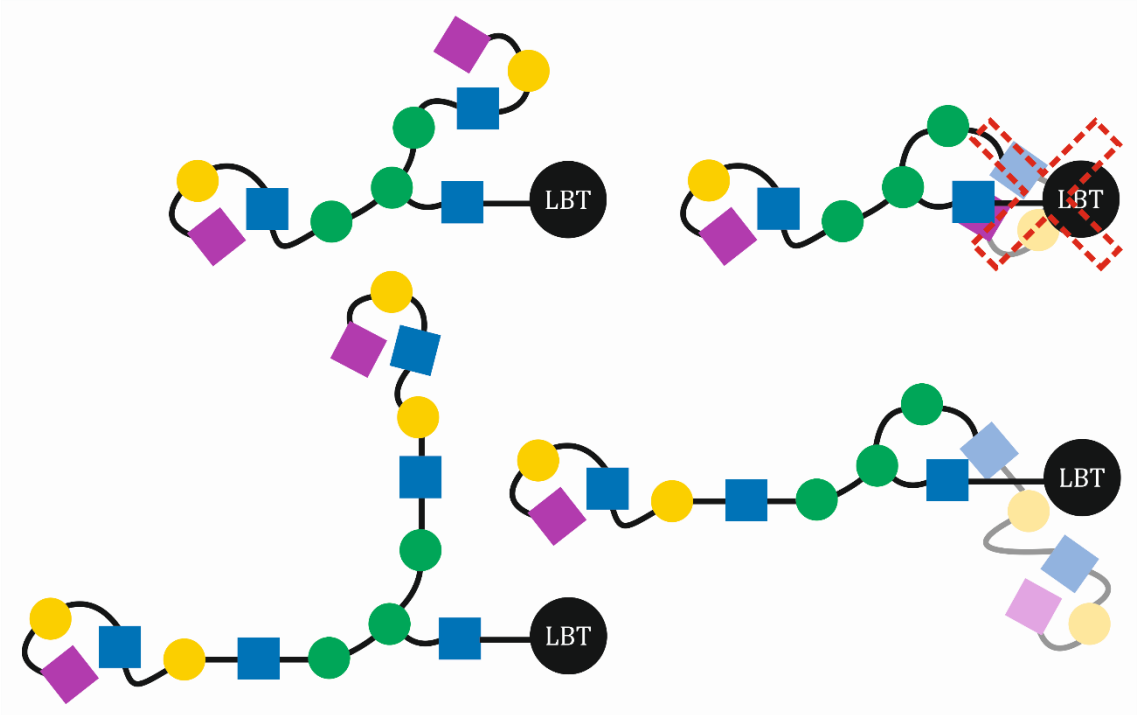


Fig 12. Schematic representation of **1** (upper schemes) and **2** (lower schemes). For the extended geometries of both N-glycans (left schemes) no steric clashes are found. In contrast, for **1** the folded geometry collapses (upper right). In the case of **2** (lower right), the additional LacNAc unit locates the terminal hairpin shape far away from the GlcNAc-LBT units, allowing the N-glycan to explore a larger conformational space and therefore to be able to adopt the folded geometry too.

As mentioned above, prior to the synthesis of N-glycans **1** and **2**, and as early step in the project, a set of three α2-6 sialylated biantennary N-glycans were synthesized (**3**, **4**, **5**), containing one, two and three LacNAc repeats at each branch respectively (figure 13). In this case, the LBT was linked to the N-glycan by a flexible Asn-Gly fragment. The flexibility introduced by this linker precluded the obtention of significant PCSs, highlighting the importance of using a rigid LBT in this type of approach.

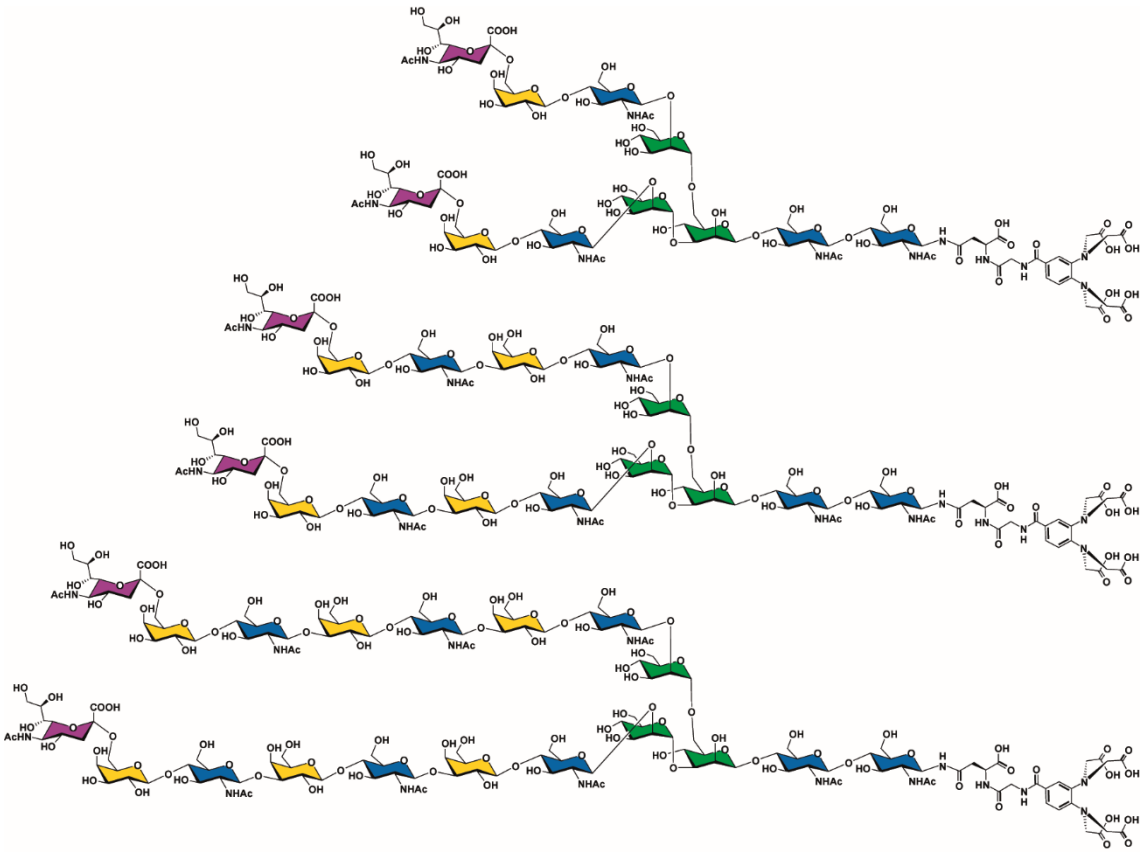


Fig 13. Structure of the N-glycans **3**, **4** and **5**.

Moreover, another N-glycan was synthesized (**1.b** as a control, similar to N-glycan **1** but without removing the GlcNAc at the reducing end (figure 14). This molecule was also studied to confirm that the removal of the first GlcNAc unit at the reducing end does not affect the conformational behavior of the glycan. The same characterization protocol as for N-glycans **1** and **2** was followed. The conformational behavior of **1.b** and **1** was the same, a two-state equilibrium composed by the extended conformers (Ψ 180°) with a major contribution of the gg rotamer (ω 60°) and a minor presence

of the gt one (ω 180°). Thus, these results confirm that the modifications introduced for **1** and **2** do not affect their conformational behavior and validate our approach.

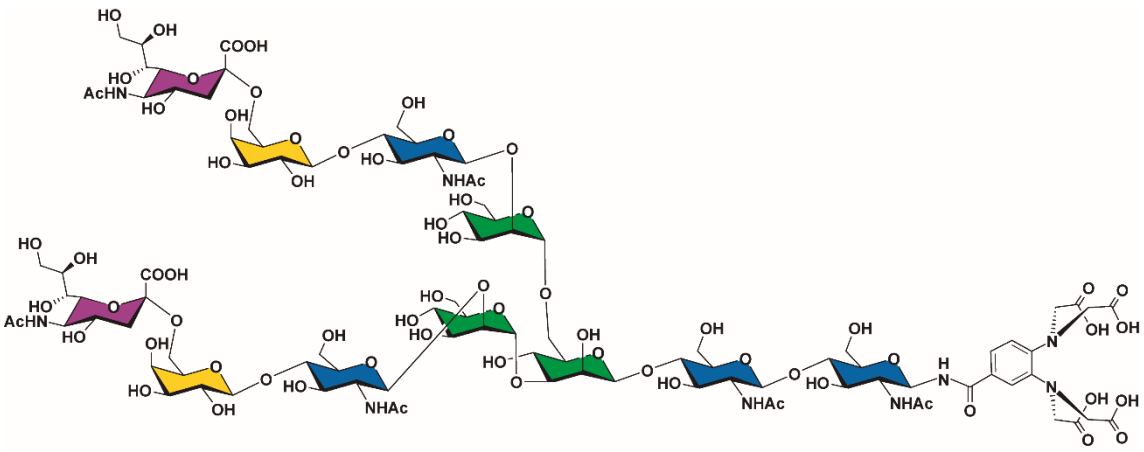


Fig 14. Structure of the N-glycan **1.b**.

4.2.3. Experimental

NMR Spectroscopy: For N-glycans **1** and **2**, NMR experiments were acquired at 310 K using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. Samples were prepared in 180 µL of deuterated TRIS buffer (D_2O, hexadeuterated TRIS, 50 mM, pH 7.8) to a final concentration of 2 mM for both derivatives in the presence of La^{3+} and Dy^{3+}, with a 1:1 carbohydrate-metal ratio. 1D ^1H NMR spectra, ^1H-^{13}C HSQC, ^1H-^{13}C HMQC-TOCSY, ^1H-^1H TOCSY and selective NOESY experiments were carried out in order to fully assign the N-glycans. The corresponding pulse sequences included in TOPSPIN Bruker software were employed. The spectra were referenced to water signal (δ=4.64 ppm for T=310 K).

The NMR assignments of N-glycans **1** and **2** are given below:

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	**Neu5Ac 6'+6**	
H_3	1.69, 2.66	40.14
H_4	3.65	68.29
H_5	3.78	52.12
H_6	3.69	72.67
H_7	3.54	68.59
H_8	3.86	71.87
H_9	3.62, 3.84	62.85
NHAc	2.01	22.30

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	**Gal 5'+5**	
H_1	4.42	103.58
H_2	3.52	70.91
H_3	3.64	72.67
H_4	3.91	68.61
H_5	3.80	73.85
H_6	3.56, 3.96	63.48

	GlcNAc 4'+4	
H_1	4.59	99.47
H_2	3.74	54.88
H_3	3.74	72.30
H_4	3.64	80.64
H_5	3.58	74.64
H_6	3.82, 3.96	60.46
NHAc	2.05	22.67

	Man α 3'	
H_1	4.94	97.12
H_2	4.11	76.53
H_3	3.90	69.55
H_4	3.50	67.49
H_5	---	---
H_6	3.62, 3.89	61.82

	Man α 3	
H_1	5.13	99.71
H_2	4.18	76.68
H_3	3.90	69.55
H_4	3.50	67.49
H_5	3.74	73.67
H_6	3.62, 3.89	61.82

	Man β 2	
H_1	4.79	100.52
H_2	4.25	70.37
H_3	3.78	80.58
H_4	3.64	74.54
H_5	---	---
H_6	3.96	66.08

	GlcNAc 1	
H_1	5.26	79.57
H_2	4.01	54.10
H_3	---	---
H_4	3.78	79.63
H_5	3.70	76.48
H_6	3.76, 3.87	60.18
NHAc	2.01	22.30

	LBT	
Arom	7.69	126.95
Arom	7.74	128.30
Arom	7.91	125.87
CH_2	3.69, 3.97	63.82

Table 4. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of N-glycan **1**.

Proton	^{1}H (δ ppm)	^{13}C (δ ppm)
	Neu5Ac 8'+8	
H_3	1.69, 2.66	40.23
H_4	3.64	68.33
H_5	3.79	52.09
H_6	3.69	72.64
H_7	3.54	68.62
H_8	3.87	71.84
H_9	3.63, 3.86	62.84
NHAc	2.02	22.23

Proton	^{1}H (δ ppm)	^{13}C (δ ppm)
	Gal 7'+7	
H_1	4.44	103.59
H_2	3.52	70.92
H_3	3.65	72.55
H_4	3.91	68.56
H_5	3.80	73.83
H_6	3.55, 3.97	63.51

	GlcNAc 6'+6	
H_1	4.72	102.71
H_2	3.78	55.12
H_3	3.77	72.31
H_4	3.65	80.46
H_5	3.59	74.49
H_6	3.82, 3.95	60.25
NHAc	2.04	22.48

	Gal 5'+5	
H_1	4.45	103.21
H_2	3.58	70.13
H_3	3.90	82.10
H_4	4.14	68.42
H_5	3.70	74.99
H_6	3.73	61.05

	GlcNAc 4'+4	
H_1	4.58	99.70
H_2	3.73	55
H_3	3.71	72.13
H_4	3.70	78.77
H_5	3.57	74.70
H_6	3.82, 3.95	60.25
NHAc	2.04	22.48

	Man α 3'	
H_1	4.94	97.12
H_2	4.11	76.53
H_3	3.89	69.63
H_4	3.50	67.50
H_5	---	---
H_6	3.62, 3.90	61.78

	Man α 3	
H_1	5.11	99.77
H_2	4.18	76.63
H_3	3.89	69.63
H_4	3.50	67.50
H_5	---	---
H_6	3.62, 3.90	61.78

	Man β 2	
H_1	4.79	100.64
H_2	4.25	70.31
H_3	3.77	80.57
H_4	3.63	74.52
H_5	---	---
H_6	3.96	65.77

	GlcNAc 1	
H_1	5.26	79.60
H_2	4.02	54.13
H_3	---	---
H_4	3.79	79.50
H_5	3.70	76.45
H_6	3.75, 3.86	60.14
NHAc	2.00	22.28

	LBT	
Arom	7.70	127.09
Arom	7.74	128.28
Arom	7.92	125.94
CH_2	3.67, 3.98	63.82

Table 5. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of N-glycan **2**.

N-glycan 1: ^{1}H NMR (600 MHz, Tris-*d*11) δ 7.91 (s, 1H), 7.74 (d, *J* = 8.6 Hz, 1H), 7.69 (d, *J* = 8.7 Hz, 1H), 5.26 (d, *J* = 9.7 Hz, 1H), 5.13 (s, 1H), 4.94 (s, 1H), 4.79 (s, 1H), 4.42 (d, *J* = 7.9 Hz, 2H), 4.25 (s, 1H), 4.18 (d, *J* = 3.6 Hz, 1H), 4.11 (d, *J* = 3.5 Hz, 1H), 4.04 – 3.45 (m, 67H), 2.66 (dd, *J* = 12.4, 4.7 Hz, 2H), 2.05 (s, 6H), 2.01 (s, 9H), 1.69 (t, *J* = 12.1 Hz, 2H).

N-glycan 2: ^{1}H NMR (600 MHz, Tris-*d*11) δ 7.92 (s, 1H), 7.74 (d, *J* = 8.6 Hz, 1H), 7.70 (d, *J* = 8.7 Hz, 1H), 5.26 (d, *J* = 9.7 Hz, 1H), 5.11 (s, 1H), 4.94 (s, 1H), 4.79 (s, 1H), 4.72 (d, *J* = 7.1 Hz, 2H), 4.45 (t, *J* = 7.2 Hz, 2H), 4.25 (d, *J* = 2.9 Hz, 1H), 4.18 (d, *J* = 3.5 Hz, 1H), 4.14 (d, *J* = 3.1 Hz, 2H), 4.10 (d, *J* = 3.4 Hz, 1H), 4.05 – 3.44 (m, 89H), 2.66 (dd, *J* = 12.4, 4.7 Hz, 2H), 2.04 (d, *J* = 3.6 Hz, 12H), 2.02 (s, 6H), 2.00 (s, 3H), 1.69 (t, *J* = 12.1 Hz, 2H).

Pseudo Contact Shift measurements and molecular modelling: PCSs were experimentally calculated as the difference in ^{1}H chemical shift measured between the ^{1}H-^{13}C HSQC spectrum acquired in the presence of a diamagnetic metal ion (La^{3+}) and the spectrum acquired in the presence of a paramagnetic metal ion (Dy^{3+}). The ^{1}H NMR chemical shifts of both N-glycans in the presence of Dy^{3+} employed to calculate the experimental PCSs are given in Table 6. Once the experimental values were obtained, the MSpin software[33,34] was used to back-calculate the expected PCSs values for the different conformations. The quality of the fitting between the experimental and calculated values is given by the Cornilescu's Q factor, which is defined by the following expression:

$$Q = \sqrt{\frac{\sum(PCS_{calc} - PCS_{exp})^2}{\sum PCS_{exp}{}^2}}$$

The components of the Δχ tensor determined by employing the experimental PCS were estimated to be $\Delta\chi_{ax} = 9.49 \cdot 10^{-32}$ m^3 and $\Delta\chi_{rh} = 5.64 \cdot 10^{-32}$ m^3 for N-glycan **1** and $\Delta\chi_{ax} = 7.85 \cdot 10^{-32}$ m^3 and $\Delta\chi_{rh} = 3.00 \cdot 10^{-32}$ m^3 in the case of N-glycan **2**.

The starting geometries of both N-glycans were downloaded from the GLYCAM web.[35] These geometries were refined using Amber 12[36] (Glycam_06h force field) in explicit water. The geometry of the phenylene diamine tetra acetic chelating unit was obtained from the reported X-ray coordinates for this moiety.[37]

	N-glycan 1
	δ (ppm)
H_5 Neu5Ac 6’	3.70
H_6 Neu5Ac 6’	3.69
H_7 Neu5Ac 6’	3.48
H_8 Neu5Ac 6’	3.80
NHAc Neu5Ac 6’	1.94
H_2 Gal 5’	3.47
H_3 Gal 5’	3.59
H_4 Gal 5’	4.02
H_5 Gal 5’	3.88
H_1 GlcNAc 4’	4.01
H_3 GlcNAc 4’	3.36
H_4 GlcNAc 4’	3.54
H_5 GlcNAc 4’	3.12
NHAc GlcNAc 4’	1.60
H_1 Man 3’	4.19
H_2 Man 3’	3.29
H_4 Man 3’	3.38
H_5 Neu5Ac 6	3.60
H_6 Neu5Ac 6	3.69
H_7 Neu5Ac 6	3.42
H_8 Neu5Ac 6	3.75
NHAc Neu5Ac 6	1.89
H_2 Gal 5	3.38
H_3 Gal 5	3.52
H_4 Gal 5	3.82
H_5 Gal 5	3.67
H_1 GlcNAc 4	4.15
H_3 GlcNAc 4	3.46
H_4 GlcNAc 4	3.40
H_5 GlcNAc 4	3.26
NHAc GlcNAc 4	1.71
H_1 Man 3	4.45
H_2 Man 3	3.65
H_4 Man 3	2.98
H_1 Man 2	3.03
H_2 Man 2	2.85
H_3 Man 2	2.77
H_4 Man 2	2.54

	N-glycan 2
	δ (ppm)
H_5 Neu5Ac 8’	3.79
H_7 Neu5Ac 8’	3.54
H_1 Gal 7’	4.53
H_2 Gal 7’	3.54
H_4 Gal 7’	3.96
H_5 Gal 7’	3.87
H_5 GlcNAc 6’	3.80
H_1 Gal 5’	4.57
H_2 Gal 5’	3.58
H_3 Gal 5’	3.94
H_4 Gal 5’	4.28
H_5 Gal 5’	3.86
H_1 GlcNAc 4’	4.03
H_3 GlcNAc 4’	3.33
H_4 GlcNAc 4’	3.66
H_5 GlcNAc 4’	3.16
NHAc GlcNAc 4’	1.60
H_1 Man 3’	4.39
H_2 Man 3’	3.41
H_5 Man 3’	2.94
H_5 Neu5Ac 8	3.76
H_7 Neu5Ac 8	3.51
H_1 Gal 7	4.41
H_2 Gal 7	3.48
H_4 Gal 7	3.89
H_5 Gal 7	3.78
H_5 GlcNAc 6	3.52
H_1 Gal 5	4.28
H_2 Gal 5	3.44
H_3 Gal 5	3.61
H_4 Gal 5	4.05
H_5 Gal 5	3.57
H_1 GlcNAc 4	4.16
H_3 GlcNAc 4	3.42
H_4 GlcNAc 4	3.46
H_5 GlcNAc 4	3.27
NHAc GlcNAc 4	1.74
H_1 Man 3	4.45
H_2 Man 3	3.66
H_5 Man 3	2.88
H_1 Man 2	3.06
H_3 Man 2	2.78
H_4 Man 2	2.55

Table 6. ^{1}H NMR chemical shifts of N-glycans **1** and **2** in the presence of Dy^{3+} employed to calculate the PCSs.

4.3. THE INTERACTION OF N-GLYCANS WITH THE HK/68 AND VIC/11 H3N2 INFLUENZA A VIRUS HEMAGGLUTININS

4.3.1. Introduction

When a strain has been stablished in the human population, HA and NA proteins go through changes to evade the immune system. However, while the antigenic drift takes place, the human immune system targets antigenic sites in and around the virus RBS. To evade this response, the viruses introduce mutations to diminish the host recognition without affecting the binding.[23] These amino acid variations provoke remarkable consequences including changes in the surface charge, the appearance of new potential glycosylation sites and variations in the binding properties.[38]

The avidity of H3N2 viruses for humans (α2-6 SLN) has decreased over the years. For instance, the viruses from 2010 do not bind 6SLN under regular conditions. The antigenic drift that occurs during the evolution of H3N2 provokes changes in the binding of the HA to the receptors, blearing the specificity for the human-type receptor.[38]

As mentioned above, recent studies using glycan microarrays[27] have shown that the HAs isolated from H3N2 viruses between 1968 and 2011 preferentially recognize extended α2-6 sialylated glycans. Furthermore, a preference for N-glycans presenting, at least, three LacNAc repetitions at each branch has been observed after 2003. Nevertheless, poor or no binding has been detected for the α2-3 linked analogues, indicating that the preference for human receptors is conserved.

The evolution of the HA from the HK/68 strain to the Vic/11 generates six more potential N-glycosylation positions in Vic/11 (13 sites) compared to those in HK/68 (7 sites).[23] Indeed, the preference for branched extended N-glycans by the later HAs was initially attributed to the increase of glycosylation points during the antigenic drift, hindering shorter receptors to properly access the RBS.[38] However, the treatment of these HAs with endoglycosidases did not show any changes in the preference for branched extended N-glycans, which was kept.[27] Therefore, the evolution of the H3N2 viruses has led to an avidity decrease for human receptors,

which affects the efficiency of infection and transmission. However, the preference for α2-6 sialylated branched extended N-glycans presenting poly LacNAc repetitions is maintained. The molecular basis of this knowledge is therefore of crucial importance for the future development of HA inhibitors.

In a collaborative project, Paulson and Woods[27] have proposed that the branched extended N-glycans could adopt a geometry that enables a simultaneous bidentate binding of the receptors with the RBS of two protomers within the same HA trimer, thus increasing the avidity. The proposed computational model was supported with the results obtained by the development of di- and tri- hemagglutinin inhibitors.[39]

In this context, we have herein addressed the study of different α2-6 sialylated branched extended N-glycans in the presence of the A/Hong Kong/1/68 (HK/68) and A/Victoria/361/11 (Vic/11) strains. The final aim is to provide experimental knowledge to obtain a detailed understanding of the molecular and structural basis of this molecular recognition event.

4.3.2. Results and discussion

The work presented herein has been carried out with three N-glycans (**2**, **4**, **5**) synthesized in the group of Prof. James Paulson at The Scripps Research Institute in San Diego, California.

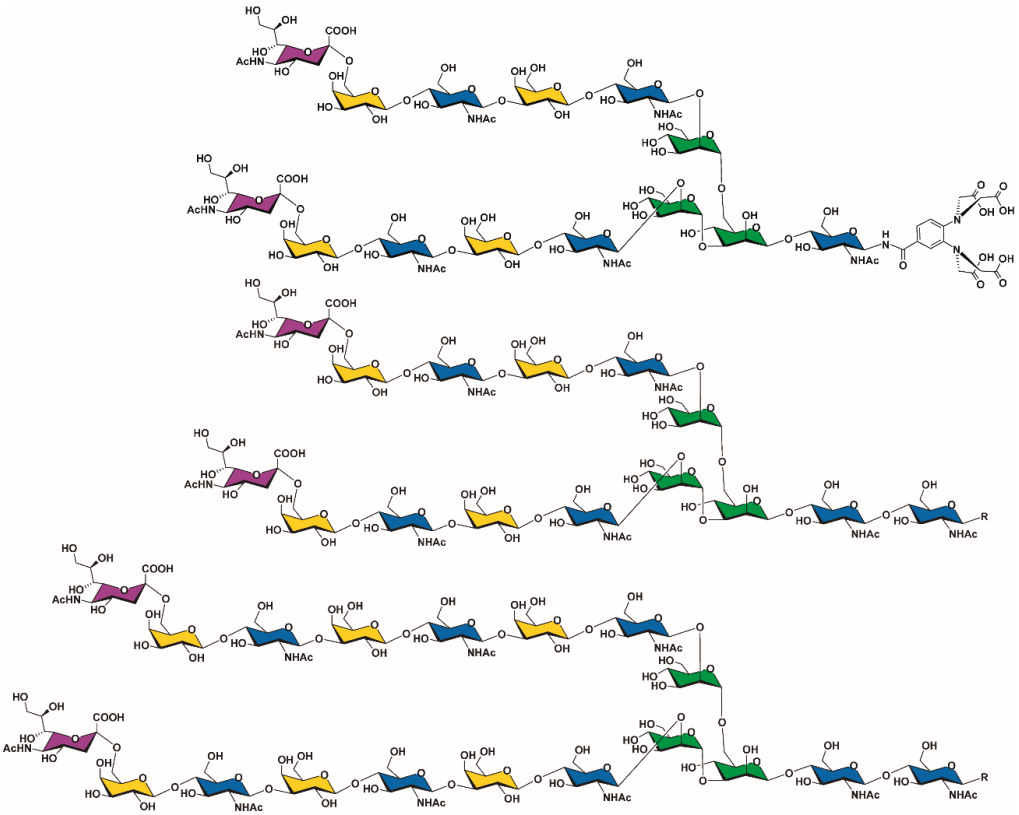


Fig 15. Schematic view of N-glycans **2**, **4** and **5**.

4.3.2.1. Interaction studies using the paramagnetic approach

The conformation of N-glycan **2**, containing two LacNAc repeats, has been described in the previous section of the chapter. Once each saccharide unit has been characterized and both branches are differenced in the NMR spectra, the interaction with the HK/68 and Vic/11 H3N2 hemagglutinins was evaluated. N-glycan **2** was chosen over N-glycan **1** based on the previous avidity assays.[27]

The interaction of **2** with the HAs was assessed using ligand-based STD NMR experiments, in agreement with the glycan microarrays studies. However, the NMR analysis permits to go deeper into the recognition details. The STD data of N-glycan **2** showed binding to the HK/68 strain involving the Neu5Ac residues, but no discrimination of the specific involvement of the different branches could be evidenced due to the chemical shift degeneracy (figure 16, left). However, the dispersion of the signals in the NMR spectra recorded under the paramagnetic conditions (sample loaded with Dy^{3+}) allowed to demonstrate the interaction of both Neu5Ac residues (8 and 8') in a non-ambiguous manner (figure 16, right). These data indicate that both arms effectively interact with the HA.

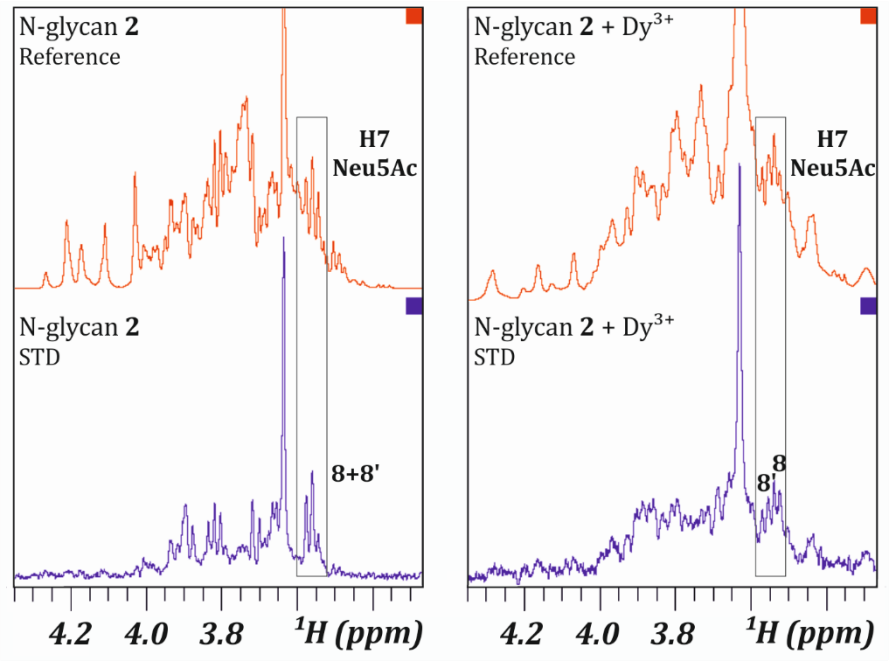


Fig 16. STD NMR experiment of N-glycan **2** (left) and loaded with Dy^{3+} (right). Signals observed in the STD spectra (in blue) show the protons that are involved in the interaction process. The H7 NMR signals of both Neu5Ac units (8 and 8') are highlighted.

The paramagnetic approach also facilitated the comparison of the interaction of **2** with both HK/68 and Vic/11 H3N2 hemagglutinins. In the absence of dysprosium, the acetyl groups of the GlcNAc units of the four lactosamine units (GlcNAc 4, 4', 6, 6') showed the same chemical shifts. However, in the presence of Dy^{3+}, a single signal for each unit is observed.

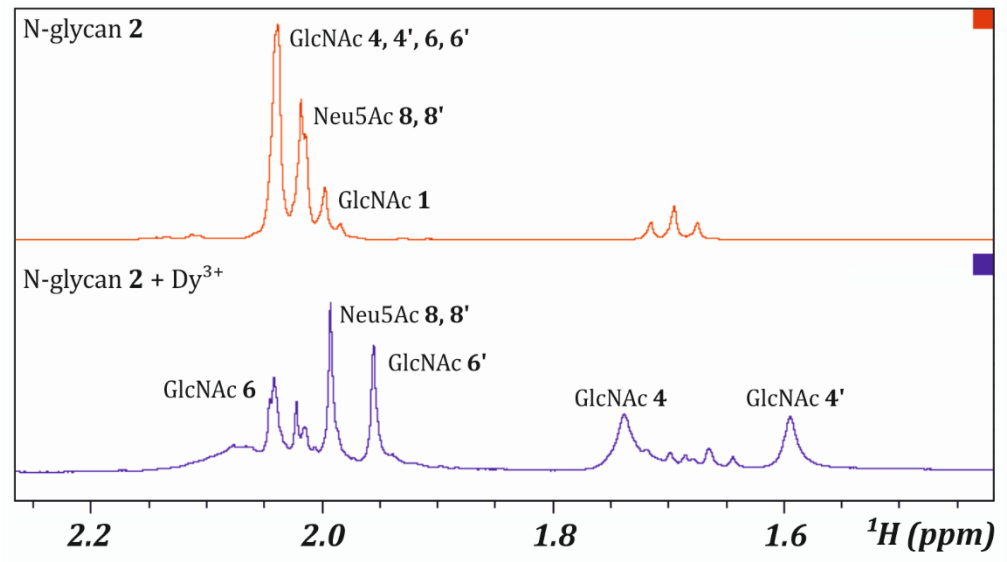


Fig 17. ^{1}H NMR spectra of N-glycan **2** in the absence (top, in red) and in the presence of Dy^{3+} (bottom, in blue). The obtention of an individual signal for each acetyl group of the GlcNAc units under paramagnetic conditions is shown.

Taking advantage of the signal dispersion in the ^{1}H NMR spectra, STD NMR experiments of **2** in the presence of both HA strains were performed and analyzed. The data showed that N-glycan **2** is recognized by both HA strains (figure 18), as expected from the arrays results.[27] However, although the data additionally showed that both branches are interacting in both cases, the STD pattern was dramatically different for both HA strains. For Vic/11, the signals corresponding to the internal carbohydrate units (GlcNAc 6 and 6') were significantly stronger, highlighting that these residues are more involved in the interaction than in the case of HK/68.

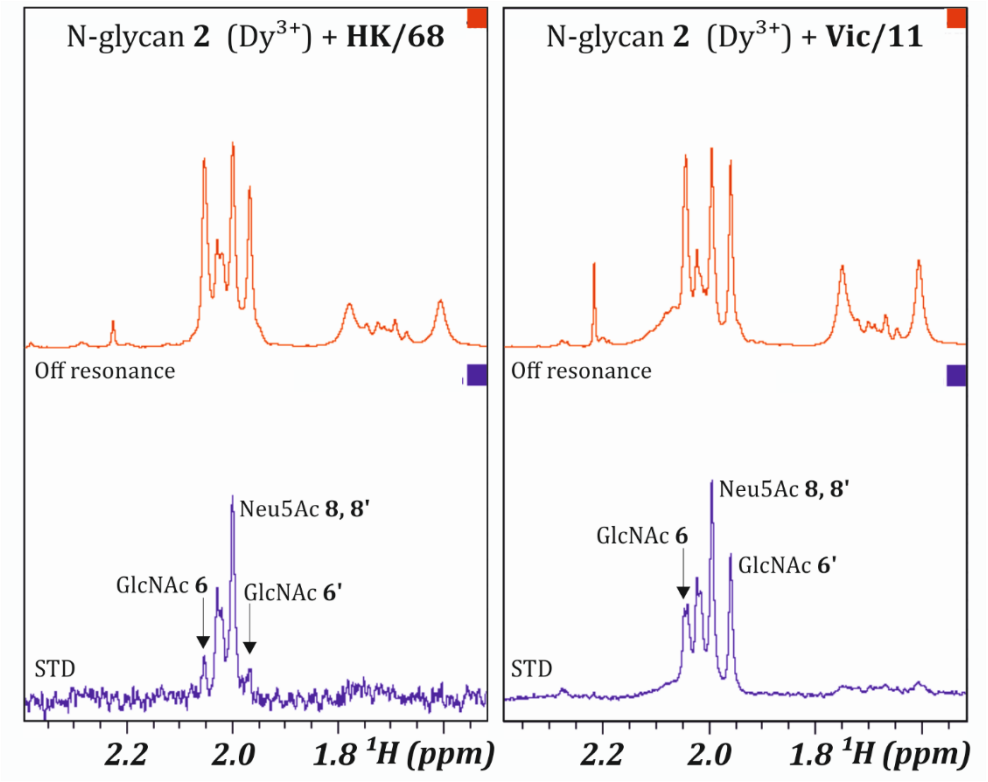


Fig 18. STD NMR experiment of N-glycan **2** loaded with Dy^{3+} in the presence of HK/68 HA strain (left) and Vic/11 HA strain (right). The STD NMR spectra (in blue) show that both branches of the N-glycan are recognized by both HA strains (STD is observed for the acetyl groups of GlcNAc 6 and 6') and a different involvement of carbohydrate units is observed for each strain.

4.3.2.2. Interaction studies between human glycans and Influenza A Virus Hemagglutinins

As mentioned above, differences in the avidity between HK/68 and Vic/11 H3N2 hemagglutinins have been reported by the combined use of glycan microarray data and molecular modelling.[27] Thus, ligand-based STD NMR experiments have been

performed herein to validate the theoretically proposed molecular recognition details between N-glycans **4** and **5** and HK/68 and Vic/11 H3N2 hemagglutinins.

4.3.2.2.1. Interaction studies of 4 and 5 with HK/68 HA

STD NMR experiments were acquired for **4** and **5** in the presence of HK/68.

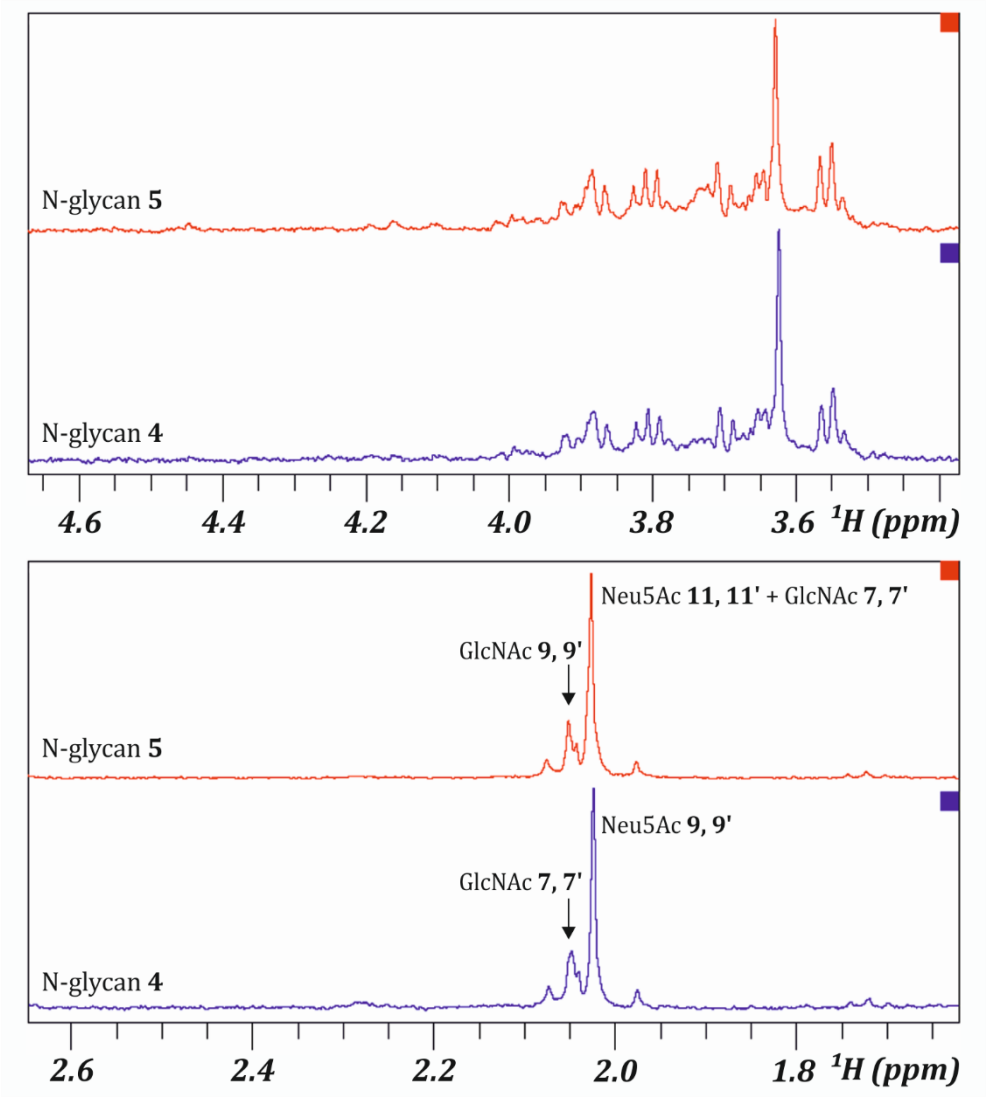


Fig 19. STD NMR experiment of **4** and **5** in the presence of HK/68 HA strain. The sugar region and acetyl expansion are shown in upper and lower panels, respectively. Key signals are labelled.

No significant differences are observed in the STD patterns of **4** and **5** in their complexes with the HK/68 hemagglutinin strain. These results point out that the presence of one extra LacNAc repeat does not stablish additional contacts with the lectin. The signals of the Neu5Ac showed the highest STD effects. Moreover, a drastic decrease of the STD intensities was observed for the internal units.

4.3.2.2.2. Interaction studies of 4 and 5 with Vic/11 HA

STD NMR experiments were also acquired for **4** and **5** in the presence of Vic/11.

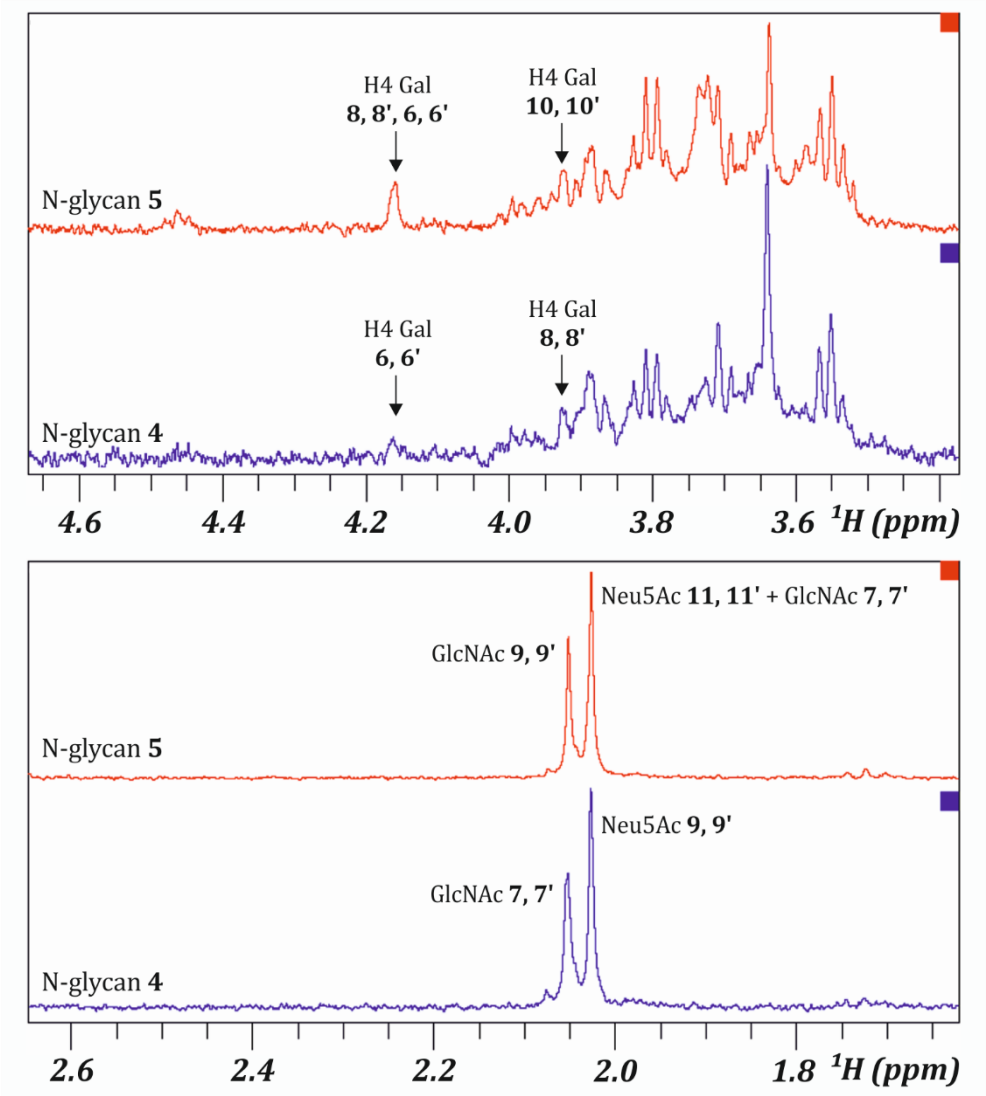


Fig 20. STD NMR experiment of **4** and **5** in the presence of Vic/11 HA strain. The sugar region and the acetyl expansion are shown in upper and lower panels, respectively. Key signals are labelled.

The STD pattern of both N-glycans was fairly similar. However, differences were observed in the relative STD intensities. The 100% of absolute STD intensity for **4** corresponded to a 10% of relative intensity, while for N-glycan **5** was significantly higher (14.3%). The H4 of the internal Gal residues showed values of 2.7% for **4** (H4 Gal 6,6'), whereas a 6% was measured for **5** (H4 6,6'+8,8'). The external residues also presented different values: 3.1% and 6.3% for **4** (H4 Gal 8,8') and **5** (H4 Gal 10,10'), respectively. In the case of Neu5Ac H7, STD values of 5.8% and 10% are observed for **4** and **5**, respectively.

Fittingly, these observations agree with the described preference of Vic/11 HA for longer chains N-glycans.

4.3.2.2.3. Comparison of the interaction of 5 with HK/68 and Vic/11 HAs

STD NMR experiments were also acquired for **5** in the presence of HK/68 and Vic/11.

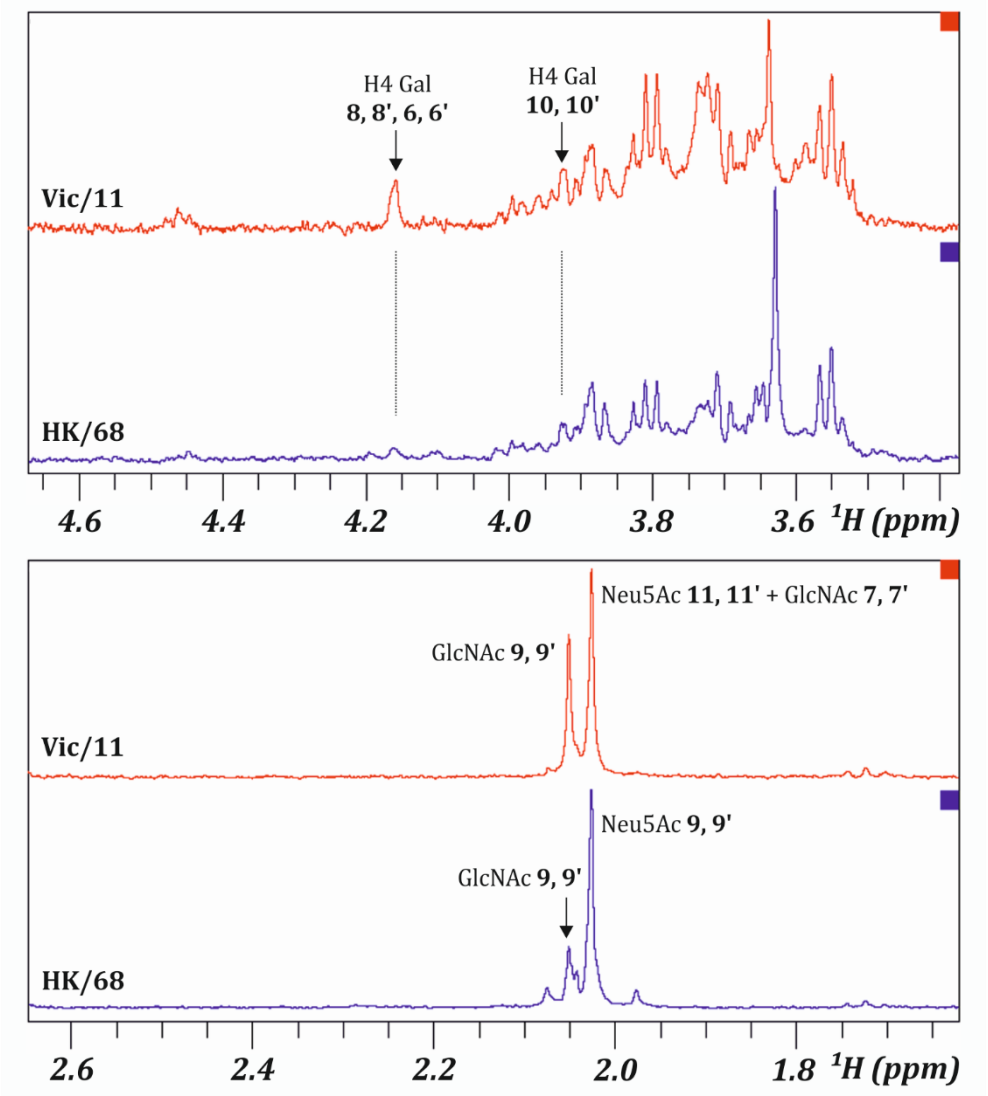


Fig 21. STD NMR experiment of **5** in the presence of Vic/11 and HK/68 HA strains. The sugar region and the acetyl expansion are shown in upper and lower panels, respectively. Key signals are labelled.

A different binding epitope is observed for **5** in the presence of Vic/11 and HK/68 HA. Additional sugar units are involved in the glycan recognition by the Vic/11 strain. STD for the external Gal (10,10') were observed for both N-glycans, whereas for internal Gal (6,6',8,8') only Vic/11 strain showed interaction. Moreover, STD for

Neu5Ac were the highest in both cases but again, GlcNAc 9,9´ displayed a higher effect in Vic/11 than in HK/68, confirming that there is a more extended epitope in Vic/11. These observations again agree again with the preference of Vic/11 HA for longer chains N-glycans.

A summary of the STD values and the epitope mapping of **4** and **5** in the presence of both Vic/11 and HK/68 strains are given below. The STD values are normalized considering the value of the highest effect as 100%.

	N-glycan **4** **Vic/11**	N-glycan **5** **Vic/11**	N-glycan **4** **HK/68**	N-glycan **5** **HK/68**
H_4 Gal internal	19	42	1	4
H_4 Gal external	22	44	13	11
H_7 Neu5Ac	40	70	24	22
NHAc GlcNAc external	44	93	12	11
NHAc Neu5Ac (+GlcNAc medium)	70	100	44	35

Table 7. STD normalized values for both **4** and **5** in the presence of Vic/11 and HK/68 HA strains.

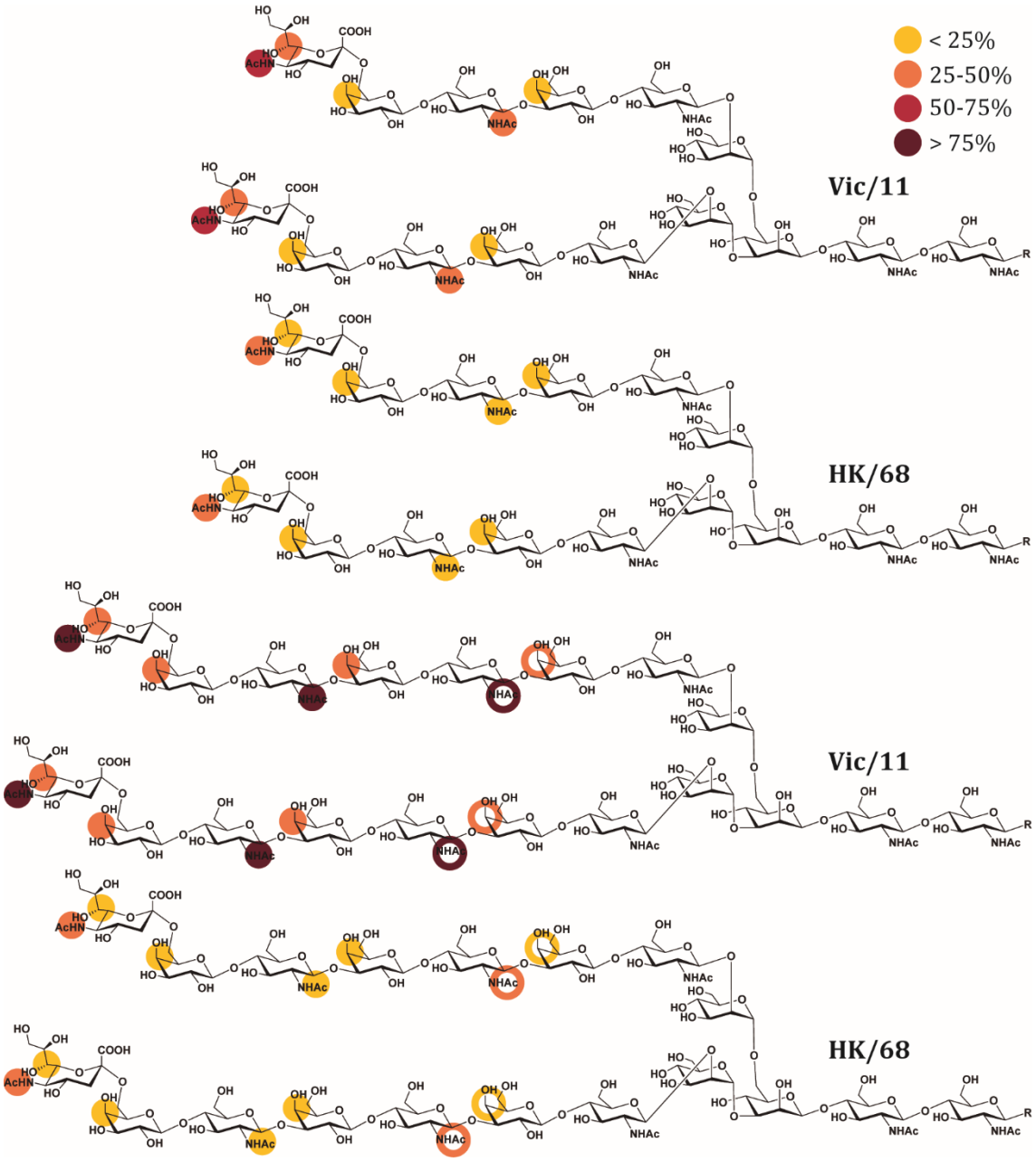


Fig 22. Schematic representation of the observed STD data for **4** and **5** (top and bottom, respectively) in the presence of Vic/11 and HK/68 HA strains. Empty circles represent overlapped proton signals.

4.3.2.2.4. The interaction model of N-glycan 5 with Vic/11 HAs

The PDB coordinates of the computational model of N-glycan **5** complexed to Vic/11 were provided by Prof. Robert J. Woods at The Complex Carbohydrate Research Center (University of Georgia). The predicted STD values for these coordinates were obtained by employing the CORCEMA package.[40]

Fittingly, the calculated STD values for the model agreed with the experimental ones, further supporting the involvement of the internal sugar units in the interaction. In particular, the STD of **5** bound to Vic/11 showed that the Ac groups of Neu5Ac (11, 11') and external GlcNAc (9, 9') moieties are involved in the interaction. In the computational model, both groups are located towards the protein surface, supporting the proposed binding mode (figure 23). The calculated STD for the previously mentioned Ac groups are 12% and 9% respectively, in full agreement with the experimental ones (14% and 9%). The Ac signal of the Neu5Ac overlays with the GlcNAc located at the second lactosamine unit (7, 7'), whose predicted STD is 1.3%. According also to the model, the external and internal Gal units (10, 10', 8, 8') are close to the protein surface, in agreement with the experimental STD.

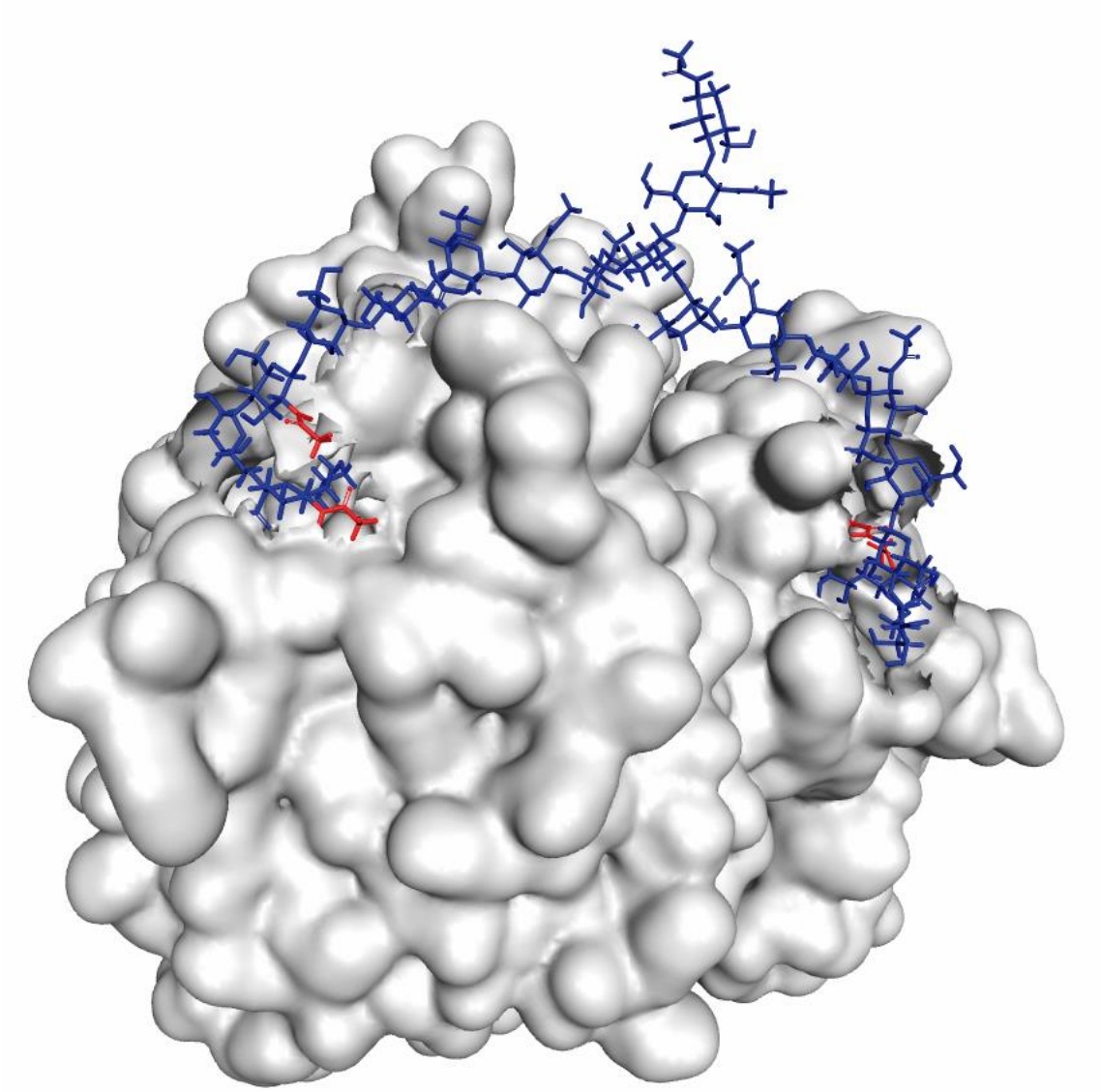


Fig 23. The computational 3D model showing the bidentate binding of **5** (blue) to Vic/11 (grey surface). The involved Ac groups of Neu5Ac (11, 11') and external GlcNAc (9, 9') are in red.

Therefore, while no significant differences have been observed in the STD pattern of **4** and **5** in complex with HK/68, differences in the relative STD intensities are detected for the interaction with Vic/11. Moreover, for N-glycan **5**, a different epitope is observed for the interaction with HK/68 and Vic/11 strains, with the second one presenting a more extended epitope.

4.3.3. Experimental

NMR Spectroscopy: For N-glycans **4** and **5**, NMR experiments were acquired at 310 K in the free state and at 293 K for the protein interaction studies, using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. Protein samples were prepared in deuterated TRIS buffer (D_2O, hexadeuterated TRIS, 50 mM, 50 mM NaCl, pH 7.8) to a final protein:ligand ratio of 1:61 (4.9 μM: 300 μM) in the case of N-glycans **2**-Dy^{3+}, **4** and **5** in the presence of HK/68 HA strain and, in the case of Vic/11 HA strain, to a final protein:ligand ratio of 1:53 (15.37 μM: 800 μM) for N-glycan **2**-Dy^{3+}, 1:67 (3.28 μM: 220 μM) for N-glycan **4** and 1:45 (6.60 μM: 300 μM) for N-glycan **5**.

1D ^1H NMR spectra, ^1H-^{13}C HSQC, ^1H-^{13}C HMQC-TOCSY, ^1H-^1H TOCSY and selective NOESY experiments were carried out to fully assign the N-glycans. The corresponding pulse sequences included in TOPSPIN Bruker software were employed. STD-NMR experiments were performed for the interaction studies. An off-resonance frequency of 100 ppm and an on-resonance frequency of -0.3 ppm were applied in the case of the N-glycans **2** (without Dy^{3+}), **4** and **5** and an on-resonance frequency of 6.5 ppm was employed in the case of Dy^{3+} sample. The saturation time was 2s. The spectra were referenced to water signal (δ=4.83 ppm for T=293K and δ=4.64 ppm for T=310K).

The NMR assignments of **4** and **5** are given below:

	^1H (δ ppm)	^{13}C (δ ppm)
Proton	**Neu5Ac 9'+9**	
H_3	1.70, 2.67	40.19
H_4	3.65	68.46
H_5	3.79	52.12
H_6	3.69	72.62
H_7	3.55	68.70
H_8	3.87	71.93
H_9	3.63, 3.86	62.88
NHAc	2.02	22.15

	^1H (δ ppm)	^{13}C (δ ppm)
Proton	**Gal 8'+8**	
H_1	4.44	103.56
H_2	3.53	70.99
H_3	3.66	72.69
H_4	3.92	68.70
H_5	3.80	73.87
H_6	3.55, 3.98	63.56

	GlcNAc 7'+7	
H_1	4.73	102.71
H_2	3.78	55.21
H_3	3.78	72.39
H_4	3.65	80.55

	Gal 6'+6	
H_1	4.46	103.26
H_2	3.59	71.21
H_3	3.72	82.19
H_4	4.14	68.45

H_5	3.60	74.55
H_6	3.83, 3.95	60.27
NHAc	2.04	22.39

H_5	3.71	75.18
H_6	3.74	61.10

	GlcNAc 5'+5	
H_1	4.57	99.69
H_2	3.73	55.06
H_3	3.71	72.24
H_4	3.70	78.83
H_5	3.57	74.84
H_6	3.83, 3.95	60.27
NHAc	2.04	22.39

	Man α 4'	
H_1	4.91	97.25
H_2	4.09	76.65
H_3	3.88	69.66
H_4	3.49	67.56
H_5	---	---
H_6	3.61, 3.89	61.78

	Man α 4	
H_1	5.11	99.67
H_2	4.17	76.67
H_3	3.88	69.66
H_4	3.49	67.56
H_5	3.73	73.50
H_6	3.61, 3.89	61.78

	Man β 3	
H_1	4.75	100.58
H_2	4.23	70.33
H_3	3.76	80.59
H_4	3.60	74.55
H_5	---	---
H_6	3.95	65.97

	GlcNAc 2	
H_1	4.60	101.38
H_2	3.78	55.21
H_3	---	---
H_4	3.72	79.68
H_5	---	---
H_6	3.73, 3.87	60.27
NHAc	1.97	22.18

	GlcNAc 1	
H_1	5.03	78.42
H_2	3.82	53.88
H_3	3.60	73.08
H_4	3.63	78.86
H_5	3.54	76.33
H_6	3.60, 3.79	60.20
NHAc	2.06	22.39

	Gly - Asn	
Gly	4.09	43.17
Asn	2.69	38.29
Asn	2.78	38.29
Asn	4.53	51.60

	LBT	
Arom	6.91	118.83
Arom	7.35	121.50
Arom	7.38	119.39
CH_2	4.19, 4.02	54.36, 54.71

Table 8. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of N-glycan **4**.

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	Neu5Ac 11'+11	
H_3	1.70, 2.67	40.22
H_4	3.65	68.48
H_5	3.79	52.13
H_6	3.70	72.58
H_7	3.55	68.73
H_8	3.88	71.92
H_9	3.63, 3.86	62.87
NHAc	2.02	22.29

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	Gal 10'+10	
H_1	4.45	103.25
H_2	3.53	70.97
H_3	3.66	72.67
H_4	3.92	68.60
H_5	3.81	73.84
H_6	3.56, 3.98	63.52

	GlcNAc 9'+9	
H_1	4.71	102.76
H_2	3.78	55.29
H_3	3.76	72.42

	Gal 8'+8	
H_1	4.45	103.25
H_2	3.59	70.11
H_3	3.72	82.18

H_4	3.65	80.51
H_5	3.58	74.63
H_6	3.82, 3.95	60.21
NHAc	2.04	22.52

H_4	4.14	68.45
H_5	3.71	75.15
H_6	3.74	61.03

	GlcNAc 7'+7	
H_1	4.71	102.76
H_2	3.78	55.29
H_3	3.76	72.42
H_4	3.71	78.61
H_5	3.58	74.63
H_6	3.82, 3.95	60.21
NHAc	2.02	22.29

	Gal 6'+6	
H_1	4.45	103.25
H_2	3.59	70.11
H_3	3.72	82.18
H_4	4.14	68.45
H_5	3.71	75.15
H_6	3.74	61.03

	GlcNAc 5'+5	
H_1	4.57	99.58
H_2	3.73	55.07
H_3	3.76	72.42
H_4	3.71	78.61
H_5	3.58	74.63
H_6	3.82, 3.95	60.21
NHAc	2.04	22.52

	Man α 4'	
H_1	4.91	97.17
H_2	4.09	76.63
H_3	3.88	69.68
H_4	3.50	67.58
H_5	---	---
H_6	3.61, 3.90	61.79

	Man α 4	
H_1	5.11	99.62
H_2	4.18	76.60
H_3	3.88	69.68
H_4	3.50	67.58
H_5	3.74	73.57
H_6	3.61, 3.90	61.79

	Man β 3	
H_1	4.75	100.55
H_2	4.23	70.32
H_3	3.76	80.63
H_4	3.62	74.63
H_5	---	---
H_6	3.95	65.92

	GlcNAc 2	
H_1	4.61	101.32
H_2	3.78	55.29
H_3	---	---
H_4	3.71	79.88
H_5	---	---
H_6	3.74, 3.88	60.23
NHAc	1.97	22.22

	GlcNAc 1	
H_1	5.04	78.30
H_2	3.82	53.94
H_3	3.61	73.05
H_4	3.64	78.90
H_5	3.56	76.31
H_6	3.61, 3.82	60.18
NHAc	2.07	22.52

	Gly - Asn	
Gly	4.09	43.17
Asn	2.69	38.29
Asn	2.78	38.29
Asn	4.53	51.60

	LBT	
Arom	6.91	118.83
Arom	7.35	121.50
Arom	7.38	119.39
CH_2	4.19, 4.02	54.36, 54.71

Table 9. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of N-glycan **5**.

N-glycan 4: ^{1}H NMR (600 MHz, Tris-*d*11) δ 7.38 (s, 1H), 7.35 (d, *J* = 8.5, 1H), 6.91 (d, *J* = 8.5, 1H), 5.11 (s, 1H), 5.03 (t, *J* = 9.6 Hz, 1H), 4.91 (s, 1H), 4.76 – 4.71(m, 3H), 4.45(dd, *J* = 9.1, 5.3 Hz, 4H), 4.23 (s, 1H), 4.20 – 4.17 (m, 5H), 4.14 (d, *J* = 5.2 Hz, 2H), 4.09 (s, 3H), 4.03 – 3.46 (m, 91H), 2.83 – 2.69 (m, 2H), 2.67 (dd, *J* = 12.4, 4.7 Hz, 2H),

2.06 (d, *J* = 2.4 Hz, 3H), 2.04 (d, *J* = 6.2 Hz, 12H), 2.02 (d, *J* = 1.9 Hz, 6H), 1.97 (s, 3H), 1.70 (t, *J* = 12.2 Hz, 2H).

N-glycan 5: ^{1}H NMR (600 MHz, Tris-*d*11) δ 7.38 (s, 1H), 7.35 (d, *J* = 8.5, 1H), 6.91 (d, *J* = 8.5, 1H), 5.11 (s, 1H), 5.04 (t, *J* = 9.6 Hz, 1H), 4.91 (s, 1H), 4.71 (dt, *J* = 16.9, 7.9 Hz, 5H), 4.49 – 4.43 (m, 6H), 4.23 (s, 1H), 4.18 (s, 5H), 4.14 (d, *J* = 3.1 Hz, 4H), 4.09 (s, 3H), 4.02 (s, 4H), 4.00 – 3.46 (m, 109H), 2.82 – 2.70 (m, 2H), 2.67 (dd, *J* = 12.4, 4.7 Hz, 2H), 2.07 (d, *J* = 2.4 Hz, 3H), 2.04 (d, *J* = 6.2 Hz, 12H), 2.02 (d, *J* = 1.9 Hz, 12H), 1.97 (s, 3H), 1.70 (t, *J* = 12.1 Hz, 2H).

4.4. SYNTHESIS OF A SIALYLATED TRILACTOSAMINE LINEAR COMPOUND AND VIC/11 HEMAGGLUTININ EXPRESSION

4.4.1. Introduction

The simultaneous bidentate binding of the receptors with the RBS of two protomers within the same HA trimer,[27] and the subsequence increase of avidity could be tested by competition assays between N-glycan **5** and its analogue **5.b**, with a single arm, a syalilated trilactosamine linear structure.

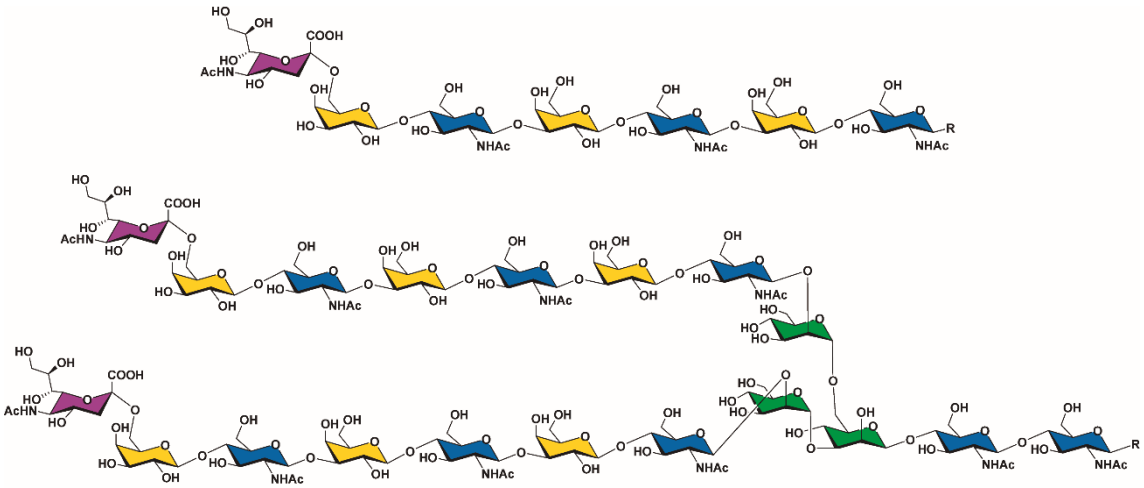


Fig 24. Structure of the proposed sialylated trilactosamine linear compound **5.b** (top) and **5** (bottom).

In this context, the synthesis of compound **5.b** was performed during my secondment period at Prof. James Paulson laboratory at The Scripps Research Institute in San Diego, California. Moreover, the expression of Vic/11 HA strain was also performed to accomplish the interaction studies.

4.4.2. Results and discussion

The synthesis of **5.b** was performed by enzymatic synthesis using a battery of glycosyltransferases in sequential steps. Glycosyltransferases are enzymes that construct specific interglycosidic linkages and their use as synthetic tools present advantages over the classical chemistry synthesis in this context. The substrate specificity and selectivity, the absence of protecting and deprotecting steps and the diminished used of solvents make the enzymatic approach highly reliable for the synthesis of glycans.[41]

The synthetic route followed for **5.b** consisted on the successive addition of Gal and GlcNAc units to the starting building block and the final addition of Neu5Ac to the terminal Gal unit through an α2-6 linkage. As reactants, glycosyl donors in the activated form (UDP- and CMP-) are required to be transferred to the acceptor scaffold by the different enzymes (figure 25).

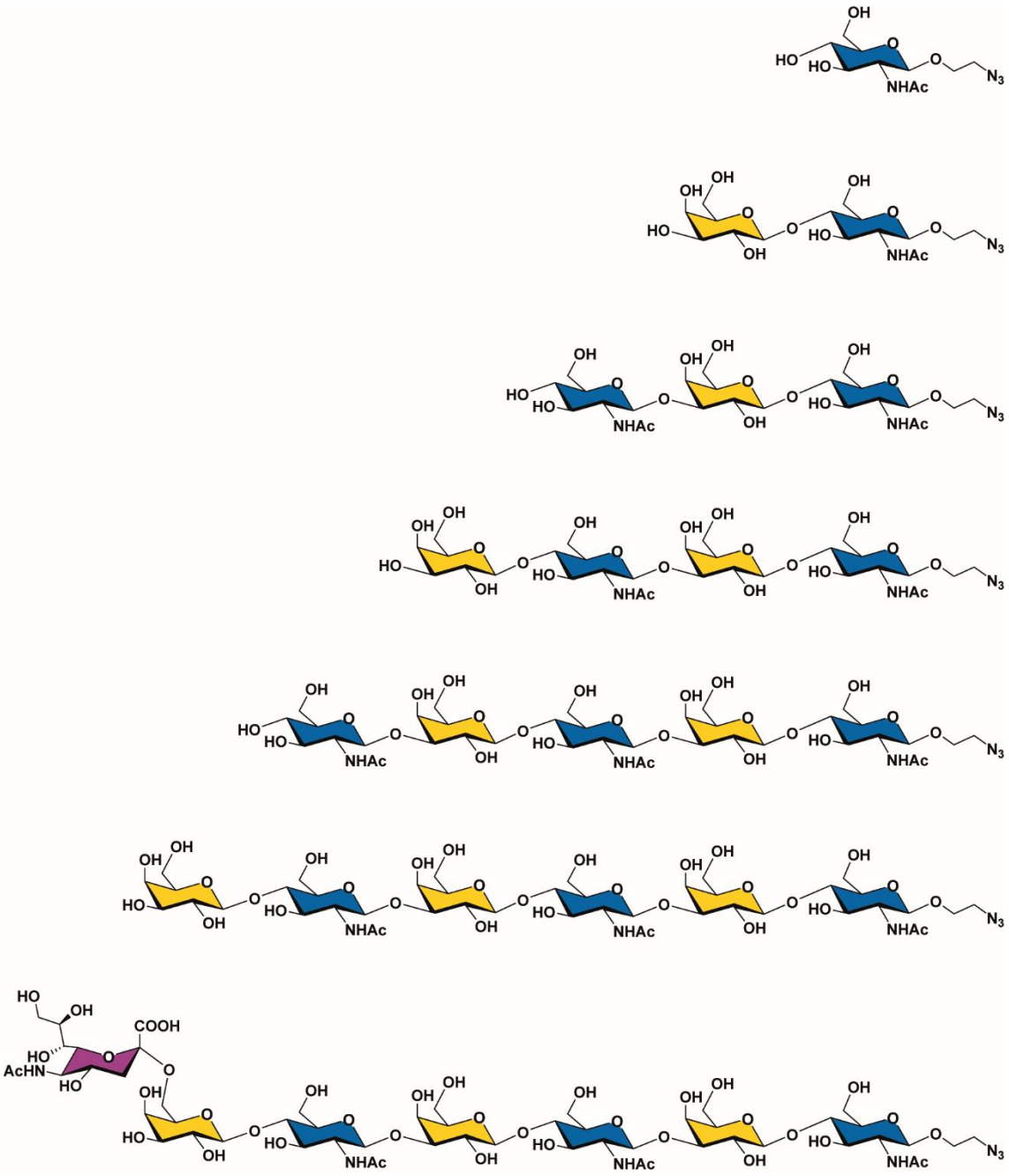


Fig 25. From top to bottom, elongation of the oligosaccharide during the synthesis by addition of one saccharide unit at the non-reducing end at each step.

For Gal addition to a GlcNAc unit at the non-reducing end, the UDP-Glc donor was employed by the chosen enzyme (β4GalT-GalE),[42] which displays epimerase and transferase activities to render a Galβ1-4GlcNAc linkage (figure 26).

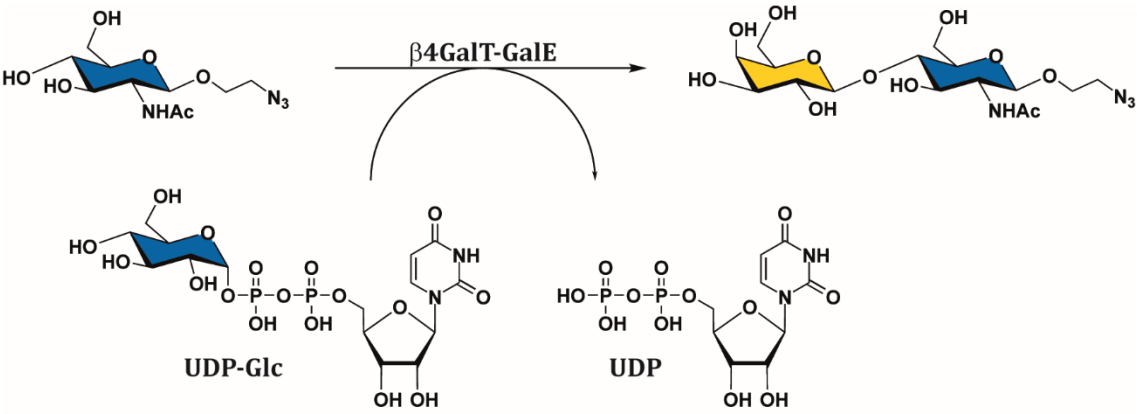


Fig 26. Gal addition to the GlcNAc unit at the non-reducing end by the action of the β4GalT-GalE.

The addition of the following GlcNAc unit to the growing oligosaccharide was performed by the GlcNAc-transferase enzyme (β3GlcNAcT),[43] to build the GlcNAcβ1-3Gal linkage (figure 27).

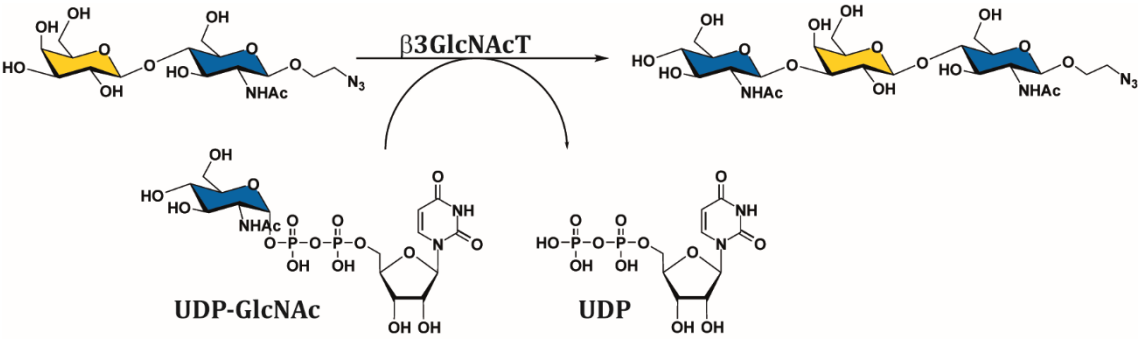


Fig 27. GlcNAc addition to the Gal unit at the non-reducing end by the action of the β3GlcNAcT.

The last step of the synthesis, after repeating twice the previous steps for obtaining the asialo-triLacNAc oligosaccharide, consisted on the addition of the Neu5Ac residue by the hST6Gal 1 sialyl transferase.[27] In this case, the required activated donor, CMP-Neu5Ac, was also enzymatically generated by using a CMP-Neu5Ac synthetase,[44] employing CTP and Neu5Ac as building blocks (figure 28).

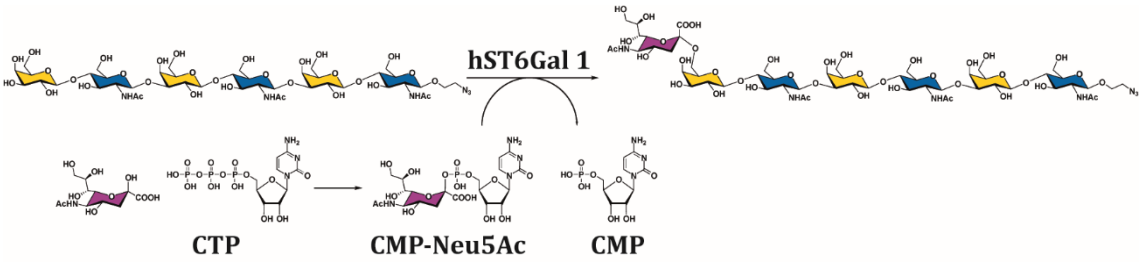


Fig 28. Neu5Ac addition to the Gal unit at the non-reducing end by the action of the hST6Gal.

Therefore, β4GalT-GalE, β3GlcNAcT and hST6Gal 1 transferases as well as CMP-Neu5Ac synthetase are required for the synthesis. However, since β3GlcNAcT and hST6Gal 1 were not initially available, their expression was also carried out during the secondment period. The experimental section contains the detailed description of both enzyme expressions.

Furthermore, the expression of Vic/11 HA strain was carried out to perform binding assays. The experimental section contains the detailed description of the expression. Competition studies with an α2-6 sialylated biantennary N-glycan containing three LacNAc repeats bearing the LBT will be performed in the near future to finish the complete description of the molecular recognition process.

4.4.3. Experimental

Gal extension: The initial mixtures containing the substrate (GlcNAc terminal glycan, 20 mM) and the donor (UDP-Glc, 2 eq of the substrate) were solved in Tris buffer (100 mM Tris, 20 mM $MnCl_2$, pH 7.5). The reactions were initiated by addition of the β4GalT-GalE enzyme (calculated based on the rxn scale for 6 hours reaction). The mixtures were incubated at 37˚C. The pH was checked after 1 hour and after that, every 3 hours. The reactions were monitored by TLC (iPrOH:NH_4OH:H_2O eluent system) and incubated until all the substrate was consumed. Once the reactions were completed, the reaction mixtures were ultra-filtered through a 4 mL, 10 KDa membrane (23˚C, 3220 rcf, 10 minutes). Then, after lyophilization, the products were re-dissolved in distilled water and load onto a PO_4^{3-} resin (ionic exchange column). In this way, the excess of UDP-Glc and the released UDP were removed. The fractions where the product was found, checked by TLC, were the subject to lyophilization. The re-dissolved product was further purified by running through a size exclusion column (P2 gel).

GlcNAc extension: The initial mixtures containing the substrate (Gal terminal glycan, 20 mM) and the donor (UDP-GlcNAc, 2 eq of the substrate) were solved in Tris buffer (100 mM Tris, 20 mM $MgCl_2$, 1 mM DTT, 25 mM KCl, pH 8). The reactions were initiated by addition of the β3GlcNAcT enzyme (calculated based on the rxn scale for 6 hours reaction). The mixtures were incubated at 37˚C. The pH was checked after 1 hour and after that, every 3 hours. The reactions were monitored by TLC ($iPrOH:NH_4OH:H_2O$ eluent system) and incubated until all the substrate was consumed. Once the reactions were completed, the reaction mixtures were ultra-filtered through a 4 mL, 10 KDa membrane (23˚C, 3220 rcf, 10 minutes). Then, after lyophilization, the products were re-dissolved in distilled water and load onto a PO_4^{3-} resin (ionic exchange column). In this way, the excess of UDP-GlcNAc and the released UDP were removed. The fractions where the product was found, checked by TLC, were the subject to lyophilization. The re-dissolved product was further purified by running through a size exclusion column (P2 gel).

Neu5Ac extension: The initial mixture containing the substrate (Gal terminal glycan, 20 mM) and the donor (CMP-Neu5Ac, 2 eq of the substrate) were solved in Tris buffer (100 mM Tris, 20 mM $MgCl_2$, pH 7.5). The reaction was initiated by addition of the hST6Gal enzyme (calculated based on the rxn scale for 6 hours reaction). The mixture was incubated at 37˚C. The pH was checked after 1 hour and after that, every 3 hours. The reaction was monitored by TLC and incubated until all the substrate was consumed. Once the reaction was completed, the reaction mixture was ultra-filtered through a 4 mL, 10 KDa membrane (23˚C, 3220 rcf, 10 minutes). Then, after lyophilization, the product was re-dissolved in distilled water and load onto a PO_4^{3-} resin (ionic exchange column). In this way, the excess of CMP-Neu5Ac and the released CMP were removed. The fractions where the product was found, checked by TLC, were the subject to lyophilization. The re-dissolved product was further purified by running through a size exclusion column (P2 gel).

β3GlcNAcT expression: After preparing and autoclave the culture medium (2XYT + 0.2% Glc), the *E. Coli* cells and Ampicillin (0.2 mL) were added to 200 mL of media and incubated overnight. 10 mL of this culture were transferred to a 1L of medium and 1 mL of Ampicillin was added. After incubation (3h), the presence of cells was

observed by UV and IPTG was added. Then, after culture incubation (21h, 220 rpm, 20°C), centrifugation (20 min, 4500 rpm, 4°C) was performed and the pellet was collected. Buffer A (Tris-HCl, NaCl, EDTA and BME) and protease inhibitor cocktail were added to the pellet, sonication was performed, followed by centrifugation (30 min, 4000 rpm, 4°C) in order to collect the periplasm presented in the supernatant. Buffer A was added to the supernatant and the sample was loaded onto an amylose column, followed by elution with buffer B (buffer A, maltose and glycerol) and collection of 1 mL fractions. 100 µL of each fraction were placed in a well plate and the absorbance was measured. Then, SDS-PAGE was performed. Finally, the enzymatic activity was calculated by performing a radioactivity assay.

hST6Gal 1 expression: Preparation of the culture medium (ultraculture medium, cholesterol, GS supplement, MSX, FBS and Penicillin) was followed by filtration for sterilization. Mammalian cells (GS-NSO) were added to the culture medium and centrifugated (5 min, 300 rpm, room temperature). 5 mL of the culture medium were added to the precipitate, transferred to a flask with 10 mL of medium and incubated. When cells had grown, hST6Gal 1 enzyme was collected (supernatant). For cells collection, 1mM EDTA in PBS was added to the flask and incubated for 10 minutes. Both hST6Gal 1 and cells were centrifugated (5 min, 300 rpm, room temperature). The supernatant contained the enzyme. The cells, precipitated after centrifugation, were culture again to further expression of the enzyme.

Vic/11 HA expression: Vic/11 was modified to remove the native viral signal peptide formed by the 10 N-terminal amino acids, and the C-terminal membrane-anchor domain formed by the final 45 amino acids. The amplified DNA product (encoding the Vic/11 ectodomain, residues 11-521) was cloned into a customized linear DNA vector for expression in mammalian tissue culture. The final Vic/11 expression fragment encoded an N-terminal CD5 signal peptide (allowing secretion of the resultant HA into the cell culture supernatant) and a C-terminal leucine zipper (GCN4) motif, to promote native trimer formation, and His8-tag for immobilized meta-affinity chromatography (IMAC) purification. Plasmid DNA harbouring the Vic/11 expression sequence was transfected into HEK293S (GnTI-/-) cells grown cell culture flasks (approximately 70% confluency) using linear PEI. After 12 hours, transfected cells were exchanged nto serum-free media and incubated for a further

72 hours at 37°C. Recombinant Vic/11 trimer was purified directly from the final cell culture supernatant by IMAC using a 1 mL HisTrap Fast Flow Crude column. Vic/11 was eluted in a gradient of PBS containing 0.5 M imidazole. Pure Vic/11 was concentrated (and buffer-exchanged to remove imidazole) in PBS.

NMR Spectroscopy: NMR experiments of **5.b** were acquired at 298 K using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe.

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	**Neu5Ac 7**	
H_3	1.64, 2.60	39.58
H_4	3.57	67.66
H_5	3.72	51.30
H_6	3.71	71.74
H_7	3.48	67.85
H_8	3.81	71.19
H_9	3.56, 3.79	62.09
NHAc	1.95	21.73

	^{1}H (δ ppm)	^{13}C (δ ppm)
Proton	**Gal 6**	
H_1	4.38	102.36
H_2	3.45	70.19
H_3	3.58	71.93
H_4	3.84	67.86
H_5	3.74	73.21
H_6	3.46, 3.92	62.79

	GlcNAc 5	
H_1	4.65	102.07
H_2	3.71	54.49
H_3	3.63	71.95
H_4	3.59	79.99
H_5	3.51	74.04
H_6	3.75, 3.88	59.47
NHAc	1.95	21.73

	Gal 4	
H_1	4.38	102.36
H_2	3.50	69.48
H_3	3.65	81.52
H_4	4.07	67.78
H_5	3.63	74.04
H_6	3.66	60.41

	GlcNAc 3	
H_1	4.65	102.07
H_2	3.71	54.49
H_3	3.63	71.95
H_4	3.63	77.79
H_5	3.51	74.04
H_6	3.75, 3.88	59.47
NHAc	1.95	21.73

	Gal 2	
H_1	4.38	102.36
H_2	3.50	69.48
H_3	3.65	81.52
H_4	4.07	67.78
H_5	3.63	74.37
H_6	3.66	60.41

	GlcNAc 1	
H_1	4.52	100.44
H_2	3.71	54.49
H_3	3.63	71.95
H_4	3.63	77.79
H_5	3.51	74.04
H_6	3.75, 3.88	59.47
NHAc	1.95	21.73

	Ethyl azide	
CH_2	3.69, 3.96	68.28
CH_2	3.35, 3.39	49.84

Table 10. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of N-glycan **5.b**.

N-glycan 5.b: ^{1}H NMR (600 MHz, D_2O) δ 4.65 (dd, *J* = 17.7, 8.0 Hz, 2H), 4.52 (d, *J* = 8.5 Hz, 1H) 4.38 (dt, *J* = 7.9, 4.8 Hz, 3H), 4.10 – 4.06 (m, 2H), 4.01 – 3.89 (m, 1H), 3.90 – 3.83 (m, 3H), 3.85 – 3.31 (m, 41H), 2.60 (dd, *J* = 12.4, 4.7 Hz, 1H), 2.01 – 1.90 (m, 12H), 1.64 (t, *J* = 12.1 Hz, 1H).

4.5. CONCLUSIONS

The conformational analysis of key N-glycans related to IAV as well as their interaction studies with two different HA strains have been performed. By applying the novel paramagnetic NMR strategy for glycans, the PCS values for every sugar unit of the target N-glycans **1** and **2** have been obtained, including those for the terminal Neu5Ac residues. Subsequent data analysis revealed that N-glycan **2** is more flexible than **1**.

Moreover, the synergic combination of paramagnetic with STD NMR data has revealed, in a non-ambiguos manner, that both branches of N-glycan **2** are invovolved in its interaction process with HK/68 and Vic/11 HA strains. The analysis of the data have also permitted to highlight the differences in the recognition features. The internal carbohydrate units are more involved in the interaction with Vic/11.

Furthermore, the interaction of N-glycans **4** and **5** with HK/68 and Vic/11 HA strains have been performed. For HK/68, no significant differences have been observed in the STD pattern of both N-glycans. The highest STD effect is observed for the terminal Neu5Ac residues with a decrease for the vicinal units. For Vic/11, the STD pattern of both N-glycans is also fairly similar, although they differ in the relative STD values. In contrast, for the interaction of N-glycan **5** with HK/68 and Vic/11 HA strains, a different binding epitope is observed. The interaction with Vic/11 takes place through a much more extended epitope.

Fittingly, the STD results support the existence of a bidentate binding mode, as previous proposed by molecular modelling.

The relationship between the bidentate binding mode and the enhance of avidity may be related to the evolution of HAs during the antigenic drift.

4.6. REFERENCES

[1] A. J. Thompson, R. P. de Vries, J. C. Paulson, *Curr. Opin. Virol.* **2019**, *34*, 117–129.

[2] J. D. Allen, T. M. Ross, *Hum. Vaccin. Immunother.* **2018**, *14*, 1840–1847.

[3] Office of the Associate Director for Communication, "Influenza (Flu) | CDC," can be found under https://www.cdc.gov/flu/, **2019**.

[4] V. N. Petrova, C. A. Russell, *Nat. Rev. Microbiol.* **2018**, *16*, 47–60.

[5] A. Varki, R. L. Schnaar, R. Schauer, *Sialic Acids and Other Nonulosonic Acids*, Cold Spring Harbor Laboratory Press, **2015**.

[6] S. J. Gamblin, J. J. Skehel, *J. Biol. Chem.* **2010**, *285*, 28403–28409.

[7] Y. Ji, Y. J. White, J. A. Hadden, O. C. Grant, R. J. Woods, *Curr. Opin. Struct. Biol.* **2017**, *44*, 219–231.

[8] "Relenza | European Medicines Agency," can be found under https://www.ema.europa.eu/en/relenza, **n.d.**

[9] "Tamiflu | European Medicines Agency," can be found under https://www.ema.europa.eu/en/medicines/human/EPAR/tamiflu, **n.d.**

[10] M. de Graaf, R. A. M. Fouchier, *EMBO J.* **2014**, *33*, 823–841.

[11] I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **1981**, *289*, 366–373.

[12] J. E. Stencel-Baerenwald, K. Reiss, D. M. Reiter, T. Stehle, T. S. Dermody, *Nat. Rev. Microbiol.* **2014**, *12*, 739–749.

[13] G. N. Rogers, J. C. Paulson, *Virology* **1983**, *127*, 361–373.

[14] J. Stevens, O. Blixt, J. C. Paulson, I. A. Wilson, *Nat. Rev. Microbiol.* **2006**, *4*, 857–864.

[15] "1918 Pandemic (H1N1 virus) | Pandemic Influenza (Flu) | CDC," can be found under https://www.cdc.gov/flu/pandemic-resources/1918-pandemic-h1n1.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Ffeatures%2F1918-flu-pandemic%2Findex.html

[16] "First Global Estimates of 2009 H1N1 Pandemic Mortality Released by CDC-Led Collaboration Error processing SSI file," can be found under https://www.cdc.gov/flu/spotlights/pandemic-global-estimates.htm

[17] "1957-1958 Pandemic (H2N2 virus) | Pandemic Influenza (Flu) | CDC," can be found under https://www.cdc.gov/flu/pandemic-resources/1957-1958-pandemic.html

[18] "1968 Pandemic (H3N2 virus) | Pandemic Influenza (Flu) | CDC," can be found under https://www.cdc.gov/flu/pandemic-resources/1968-pandemic.html

[19] L. A. Reperant, T. Kuiken, A. D. M. E. Osterhaus, *Vaccine* **2012**, *30*, 4419–4434.

[20] E. Sorrell, E. Schrauwen, M. Linster, M. De Graaf, S. Herfst, R. Fouchier, *Curr. Opin. Virol.* **2011**, *1*, 635–642.

[21] L. Glaser, J. Stevens, D. Zamarin, I. A. Wilson, A. García-Sastre, T. M. Tumpey, C. F. Basler, J. K. Taubenberger, P. Palese, *J. Virol.* **2005**, *79*, 11533–6.

[22] T. M. Tumpey, T. R. Maines, N. Van Hoeven, L. Glaser, A. Solórzano, C. Pappas, N. J. Cox, D. E. Swayne, P. Palese, J. M. Katz, et al., *Science* **2007**, *315*, 655–659.

[23] H. Yang, P. J. Carney, J. C. Chang, Z. Guo, J. M. Villanueva, J. Stevens, *Virology* **2015**, *477*, 18–31.

[24] G. N. Rogers, J. C. Paulson, R. S. Daniels, J. J. Skehel, I. A. Wilson, D. C. Wiley, *Nature* **1983**, *304*, 76–78.
[25] R. J. Connor, Y. Kawaoka, R. G. Webster, J. C. Paulson, *Virology* **1994**, *205*, 17–23.
[26] A. Chandrasekaran, A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, R. Sasisekharan, *Nat. Biotechnol.* **2008**, *26*, 107–113.
[27] W. Peng, R. P. de Vries, O. C. Grant, A. J. Thompson, R. McBride, B. Tsogtbaatar, P. S. Lee, N. Razi, I. A. Wilson, R. J. Woods, et al., *Cell Host Microbe* **2017**, *21*, 23–34.
[28] A. Canales, A. Mallagaray, J. Pérez-Castells, I. Boos, C. Unverzagt, S. André, H.-J. Gabius, F. J. Cañada, J. Jiménez-Barbero, *Angew. Chemie Int. Ed.* **2013**, *52*, 13789–13793.
[29] A. Canales, I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, G. Domínguez, F. J. Cañada, J. Pérez-Castells, D. Häussinger, et al., *Angew. Chemie Int. Ed.* **2017**, *56*, 14987–14991.
[30] B. Fernández de Toro, W. Peng, A. J. Thompson, G. Domínguez, F. J. Cañada, J. Pérez-Castells, J. C. Paulson, J. Jiménez-Barbero, Á. Canales, *Angew. Chemie Int. Ed.* **2018**, *57*, 15051–15055.
[31] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, et al., *Glycobiology* **2015**, *25*, 1323–1324.
[32] C. M. Nycholat, R. McBride, D. C. Ekiert, R. Xu, J. Rangarajan, W. Peng, N. Razi, M. Gilbert, W. Wakarchuk, I. A. Wilson, et al., *Angew. Chemie Int. Ed.* **2012**, *51*, 4860–4863.
[33] A. Navarro-Vázquez, *Magn. Reson. Chem.* **2012**, *50*, S73–S79.
[34] "Mestrelab Research S.L. - Analytical Chemistry Software Solutions," can be found under https://mestrelab.com/
[35] "GLYCAM," can be found under http://glycam.org/
[36] T. A. D. D. A. Case, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, A. W. G. J. Swails, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, S. R. B. X. Wu, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 12, University of California, San Francisco, **2012**.
[37] M. Mizuno, S. Funahashi, N. Nakasuka, M. Tanaka, *Inorg. Chem.* **1991**, *30*, 1550–1553.
[38] Y. P. Lin, X. Xiong, S. A. Wharton, S. R. Martin, P. J. Coombs, S. G. Vachieri, E. Christodoulou, P. A. Walker, J. Liu, J. J. Skehel, et al., *Proc. Natl. Acad. Sci.* **2012**, *109*, 21474–21479.
[39] W. Lu, W. Du, V. J. Somovilla, G. Yu, D. Haksar, E. de Vries, G.-J. Boons, R. P. de Vries, C. A. M. de Haan, R. J. Pieters, *J. Med. Chem.* **2019**, *62*, 6398–6404.
[40] V. Jayalakshmi, N. R. Krishna, *J. Magn. Reson.* **2002**, *155*, 106–118.
[41] H. S. Overkleeft, P. H. Seeberger, *Chemoenzymatic Synthesis of Glycans and Glycoconjugates*, Cold Spring Harbor Laboratory Press, **2015**.
[42] O. Blixt, J. Brown, M. J. Schur, W. Wakarchuk, J. C. Paulson, *J. Org. Chem.* **2001**, *66*, 2442–2448.

[43] S. M. Logan, E. Altman, O. Mykytczuk, J.-R. Brisson, V. Chandan, F. St. Michael, A. Masson, S. Leclerc, K. Hiratsuka, N. Smirnova, et al., *Glycobiology* **2005**, *15*, 721–733.
[44] M.-F. Karwaski, W. W. Wakarchuk, M. Gilbert, *Protein Expr. Purif.* **2002**, *25*, 237–40.

CHAPTER 5

CONCLUSIONS

In this Thesis, different carbohydrate-protein systems of biological relevance have been studied using a combination of Nuclear Magnetic Resonance spectroscopy with other experimental and theoretical techniques.

- A set of four sulfoxide disaccharides have been studied. The conformational analysis of the acetylated and deacetylated derivatives by NMR combined with computational methods has revealed a different conformational behavior depending on the stereochemistry at the sulfoxide centre. Furthermore, although both deacetylated derivatives act as moderate *E. coli* β-galactosidase inhibitors, the inhibition potency also depends on the stereochemistry at the sulfur atom. Despite the inhibition activity, both sulfoxides are hydrolyzed, albeit at very different rates. The high flexibility presented by **2*S*** plays an important role in the molecular recognition process and a conformational selection process takes place. The combination of NMR experimental data with modeling procedures has allowed explaining the experimental observations.

- The conformational analysis of key N-glycans receptors of the Influenza A virus has been performed, as well as the study of their interactions with the HK/68 and Vic/11 HA strains. The application of a novel paramagnetic NMR strategy for N-glycans has permitted to deduce the preferred geometries of these molecules in solution. In particular, the PCS values for every sugar unit of the target N-glycans **1** and **2** have been obtained and compared to the expected values for different possible three-dimensional shapes. The subsequent extensive data analysis revealed that N-glycan **2** can explore a wider conformational space than **1**.

  Moreover, the synergic combination of paramagnetic and STD NMR-based protocols has revealed that both branches of N-glycan **2** are involved in its interaction process with both HA strains. The analysis of the data has also permitted to highlight key differences in their recognition features.

Furthermore, the interactions of N-glycans **4** and **5** with both strains have been studied. While for HK/68 HA strain no differences are observed in the STD pattern of both N-glycans, in the case of Vic/11, although the STD pattern is fairly similar, the STD intensities significantly differ. In particular, N-glycan **5** shows a drastically different interaction mode with both strains, displaying a much more extended binding epitope in the case of Vic/11.

APPENDIX

GLYCOSIDASES FROM FUNGI: β-XYLOSIDASE AND β -GLUCOSIDASE FROM TALAROMYCES AMESTOLKIAE. A VIEW BY NMR.

6.1. INTRODUCTION

Glycosyl hydrolases (GH), also known as glycosidases and glycoside hydrolases, are one of the five enzyme classes among CAZymes that catalyze the cleavage of a glycosidic bond.[1] Although glycosidic linkages are considered between the most stable in nature, the GH enhance up to 10^{17} the hydrolysis rate compared to the spontaneous hydrolysis. Hence, glycosidases are considered one of the most efficient catalyst available.[2] In the CAZy database, glycosidases from all kingdoms are organized, based on amino acid sequence similarities, in more than 160 families amounting several thousands of different sequences.

Two criteria are used to classify GH from the mechanistic point of view. First, depending on the position of the hydrolyzed glycosidic bond, GH are distinguished in endo- and exo-glycosides (internal residue linkages and glycosidic linkages at the terminal position of the chain, respectively). The second criterium is based on the action mechanism of the enzymes, considering the configuration of the anomeric position of the reaction product. Therefore, GH are classified as inverting or retaining glycosidases. Both catalytic mechanisms (figure 1) were first described by Koshland[3] and reinforced later by more experimental evidences[4-6]. In most cases (either inverting and retaining enzymes), two catalytic amino acid residues, a general acid (proton donor) and a nucleophile/base, most frequently a side chain with carboxyl groups (Glu or Asp), are involved in the catalysis.

In the case of inverting glycosidases, one amino acid acts as an acid catalyst (protonated neutral carboxylic acid) and the other amino acid acts as a base catalyst (anionic carboxylate). When a water molecule enters into the active site, already loaded with the glycoside substrate, the carboxylate deprotonates it, enabling the attack of the activated water molecule to the anomeric position through the opposite face to the glycosidic bond. This nucleophilic attack is assisted by the protonated carboxylic acid, which protonates the inter glycosidic oxygen, weakening the linkage to the anomeric carbon. This mechanism occurs through one transition state, which releases a product with inverted configuration (figure 1A).

In the case of retaining enzymes, two consecutive inversions of the configuration take places. In the first step, the carboxylate, which acts as a nucleophile, attacks the anomeric position through the opposite face. The acid/base residue assists the catalysis by protonating the interglycosidic oxygen. Therefore, the first step renders the formation of a covalent glycosyl-enzyme intermediate with inverted anomeric configuration. The subsequent and last step consists on the deprotonation of a water molecule by the acid/base residue, which attacks the anomeric position through the opposite face of the linkage to the enzyme. After two transition states, this mechanism with two sequential inversions, renders a product with the same anomeric configuration that the substrate (figure 1B)[7,8]

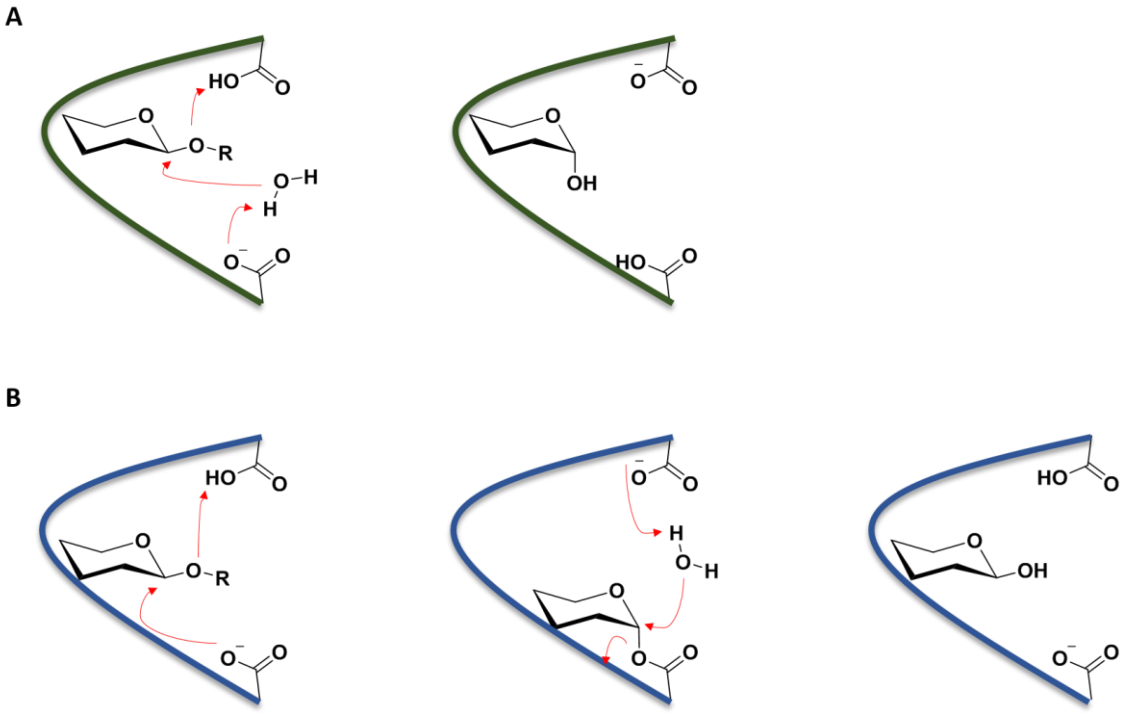


Figure 1. Mechanism of inverting (A) and retaining (B) glycosyl hydrolases.

Although the mayor activity of glycosyl hydrolases consists on the hydrolysis of the glycosidic linkages, the ability of these enzymes for also synthetizing them is called transglycosylation[9]. The generation of a glycosyl-enzyme intermediate in GHs with a retaining mechanism makes possible that, in the presence of acceptors different from water, new glycoside derivatives are formed by transferring the glycosyl residue of the intermediate to another compound (figure 2)[10,11]. The ability of GHs for catalyzing transglycosylation reactions has prompted up these enzymes as a

powerful alternative to classical organic synthesis to obtain stereo and regioselective oligosaccharides and glycoconjugates, avoiding numerous synthetic steps[12].

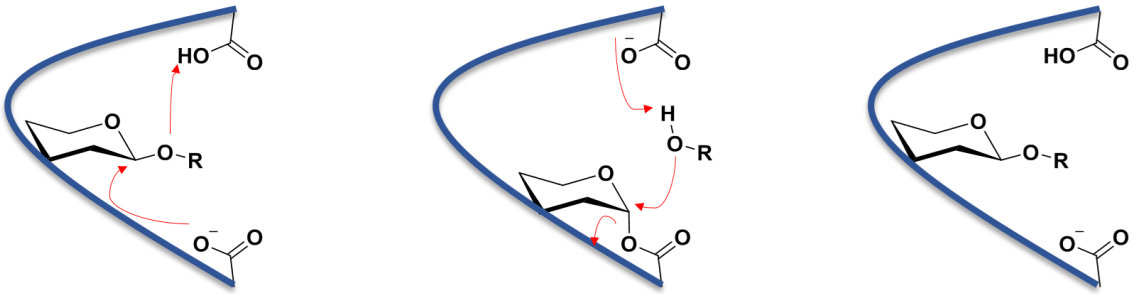


Figure 2. Mechanism of transglycosylation.

An example of the transglycosylation activity used in the clinical field is the synthesis of a lactose analogous (galactosyl-β(1-4)-xylose) to be used *in vivo* for the diagnose of lactase deficiency. The *E. coli* β-galactosidase is used in the presence of *o*NP-galactose as a substrate and a high concentration of xylose as acceptor to favour the transglycosylation[13,14]. Moreover, some types of oligosaccharides (as cellooligosaccharides, galactooligosaccharides and xylooligosaccharides) are used as prebiotics in the food industry[15].

However, despite the advantages that GHs present in the synthesis field, some weaknesses are also displayed. The low yields obtained and the related possibility of the hydrolysis of the product, which can be even a better substrate for the enzyme that the donor employed in the transglycosylation, opened the field to mutation strategies to overcome these problems[16]. Two approaches can be considered: glycosynthases[17,18] and thioglycoligases[19]. In the first case (figure 3a), the amino acid responsible of the nucleophilic attack is substituted by an inert residue, disenabling the catalysis of the hydrolysis. An activated donor mimicking the glycosyl-enzyme intermediate (i.e fluorine-glycosides) is needed to overcome the first reaction step and act as glycosyl acceptor. Then, this substrate is attacked by the acceptor with the assistance of the acid/base amino acid. The resulting product, now a non-activated glycoside, cannot be hydrolyzed by the mutated enzyme.

Regarding the thioglycoligases, they are obtained by substitution of the acid/base amino acid (figure 3B). Therefore, a donor with a good leaving group (not dependent

of acid catalysis) is necessary (i.e nitrophenyl-glycosides) to obtain the covalent glycosyl-enzyme intermediate. Then, the nucleophile attack by a strong nucleophile acceptor (i.e thiol nucleophiles) occurs.

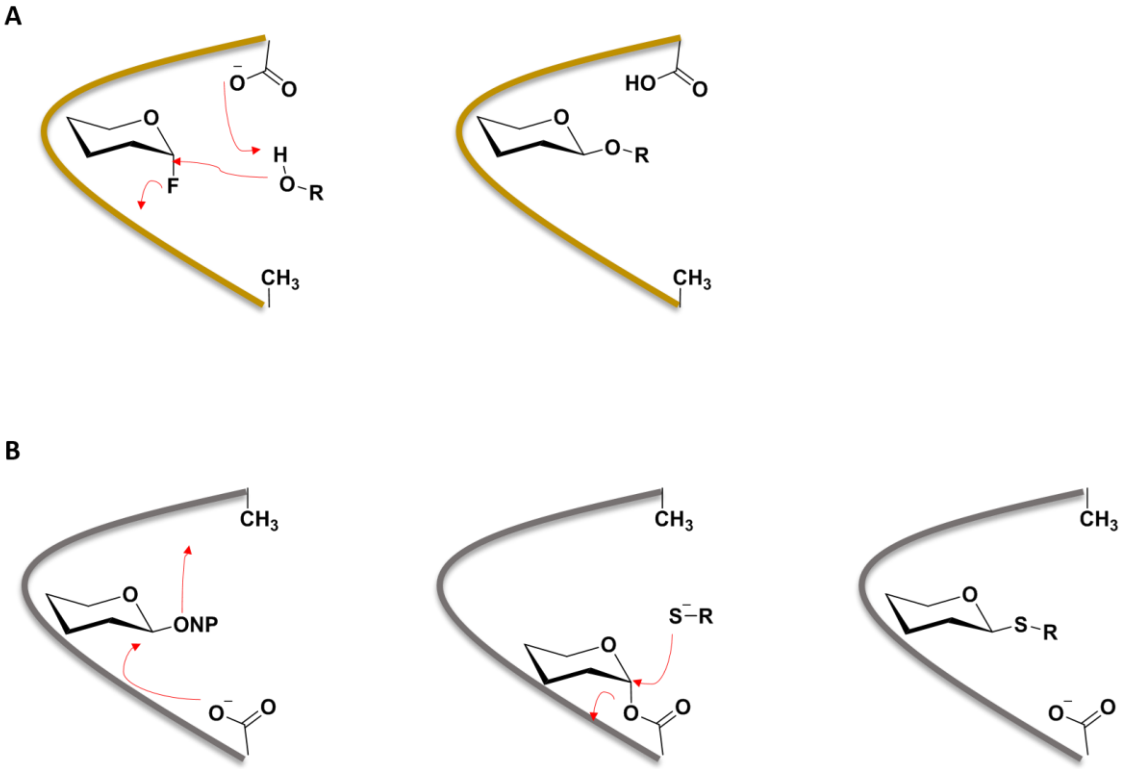


Figure 3. Mechanism of glycosyl hydrolase mutants: glycosynthases (A) and thioglycoligases (B).

In this context, the work described in this chapter has been focused on the study of two native GH enzymes and their transglycosylation activity by characterizing the products, and the mutants of these enzymes with glycosynthase and thioglycoligase activities.

The work presented herein has been carried out in collaboration with the group of Prof. María Jesús Martínez Hernández at the Biological Research Center (CIB-CSIC) in Madrid who has identified, isolated, characterized and mutated different glycosidase enzymes present in wood-degrading fungi. The application of NMR-based strategies has allowed to mechanistically characterize those enzymes and to determine the structure of their reaction products when used to catalyze transglycosylation reactions.

6.2. β-XYLOSIDASE FROM *TALAROMYCES AMESTOLKIAE*

6.2.1. Introduction

Plant biomass is the main renewable resource in biosphere, composed of two polysaccharides, cellulose and hemicellulose. Xylans are the most abundant type of hemicellulose[20], whose conversion into xylose is of great interest in the obtention of second-generation bioethanol, and a source of antioxidants, biomaterials and xylosides[21]. Therefore, the identification of new β-xylosidases is needed.

BxTW1 is a GH3 β-xylosidase from the hemicellulolytic fungus *Talaromyces amestolkiae*.[22] The hydrolytic mechanism, the substrate specificity and the transxylosylation activity of the enzyme have been studied with the help of NMR experiments.[22-24] Moreover, the analysis of the products obtained by the rBxTW1-E495A thioglycosynthase type mutant has been performed.

6.2.2. Results and discussion

6.2.2.1. Native BxTW1

6.2.2.1.1. Hydrolytic mechanism

The hydrolytic mechanism, inverting or retaining, of native BxTW1, extracted from its producing fungi, was elucidated by following the ^{1}H NMR signals of spectra recorded at different reaction times (figure 4), and allowed to determine the anomeric configuration of the xylose formed by hydrolysis of *p*NPX. The product of the enzymatic reaction was β-xylose, which spontaneously slowly isomerizes up to equilibration with the α-anomer, pointing out that BxTW1 hydrolyses the substrates by retaining the anomeric configuration.

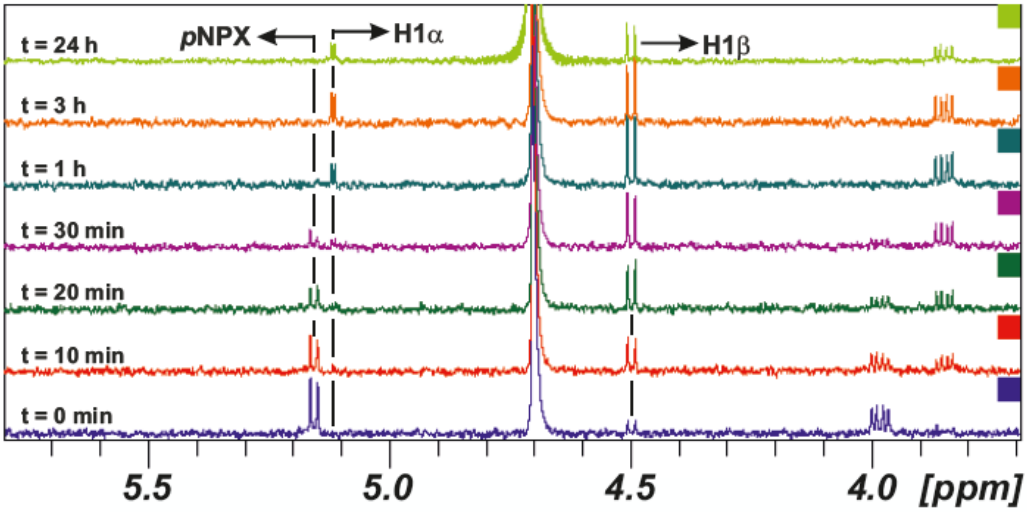


Figure 4. ^{1}H NMR spectra recorded at different times in the course of the enzymatic hydrolysis. The anomeric position of the released β-xylose is shown. Signals of α-xylose are also detected at longer times due to the spontaneous mutarotation up to equilibrium of both conformers.

6.2.2.1.2. Substrate specificity

The enzyme is able to hydrolyse xylooligosaccharides of different lengths[22]. To check if there is any length preference by BxTW1, the selectivity for xylotriose or xylobiose was studied by ^{1}H NMR following directly in the NMR tube a reaction started with the trisaccharide as substrate (figure 5).

First, the hydrolysis of xylotriose rendered xylose and xylobiose, which was also hydrolyzed to xylose. The amounts of each compound were obtained by integrating characteristic signals from the ^{1}H NMR spectra recorded at different reaction time: xylose was analyzed from H_5 (δ 3.81), xylotriose from H_5' (δ 4.04), and xylobiose (no specific signal due to overlapping) by substracting xylotriose H_5' (δ 4.04) from the signal (δ 3.90) that results from the overlapping of xylobiose and xylotriose. The evolution curves, where some accumulation of the disaccharide was observed (figure 5d), confirmed the preference of BxTW1 for xylotriose over the xylobiose, pointing out an extended binding site in the enzyme in accordance with the polymeric nature of xylan substrates.

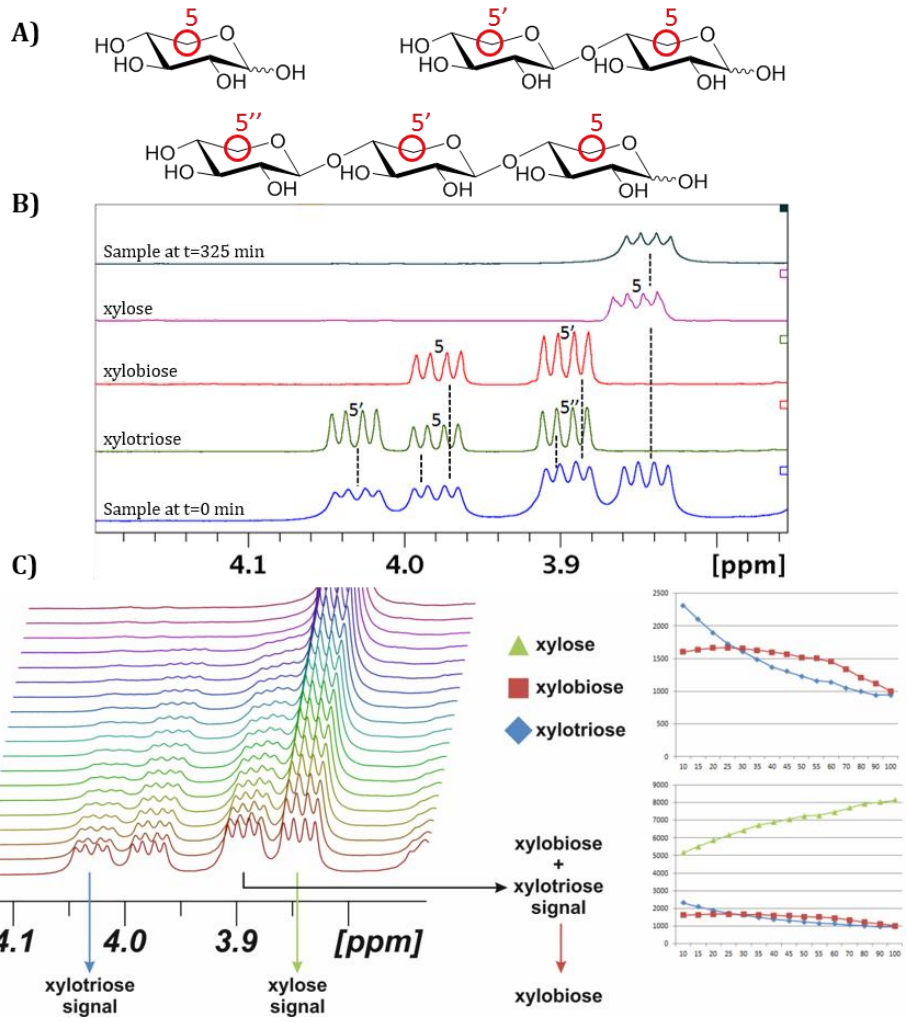


Figure 5. A) Structure of xylose, xylobiose and xylotriose. The protons corresponding to the characteristic signals used for the evolution curves are highlighted. B) Comparison of the ^{1}H 1D NMR spectra of the reaction at time 0 (blue) and when the reaction is completed (dark green), with commercial xylose (purple), xylobiose (red) and xylotriose (light green). C) ^{1}H NMR spectra of xylotriose consumption by BxTW1 in the course of the enzymatic hydrolysis. Evolution of xylotriose (blue), xylobiose (red) and xylose (green) concentrations.

6.2.2.1.3. Transxylosylation activity

The transxyloxylation activity of BxTW1 was observed in preliminary experiments using xylobiose as unique substrate. The observation of higher order oligosaccharides indicates that xylobiose itself could act as both donor and acceptor. Diffusion Ordered SpectroscopY (DOSY-NMR),[25] a technique that can distinguish oligosaccharides of different sizes, was applied. Three saccharides of different molecular weight were observed by recording the DOSY spectra of the reaction mixture (figure 6A). A pseudo 3D DOSY-TOCSY experiment was employed to identify

the saccharides presented in the reaction mixture, xylose, xylobiose and xylotriose (figure 6B).

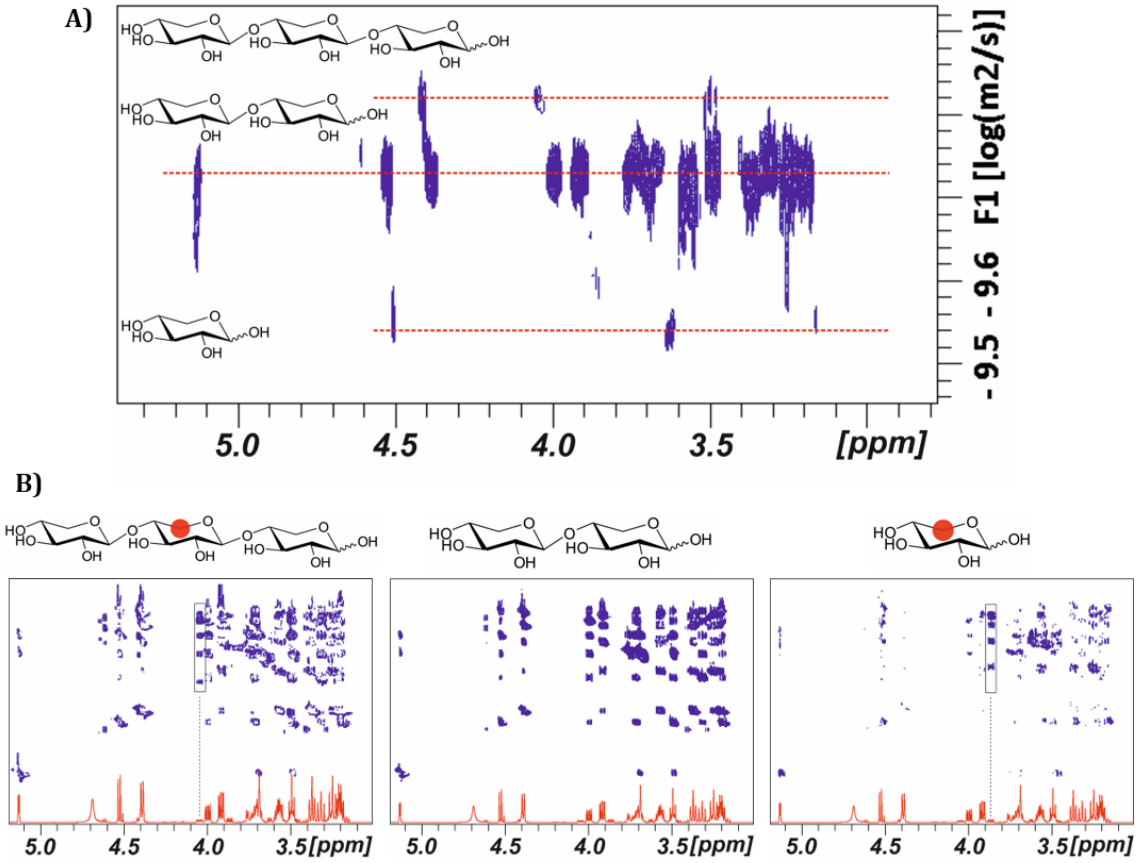


Figure 6. A) DOSY NMR spectra of the reaction mixture. B) The TOCSY planes extracted for each saccharide from the 3D DOSY-TOCSY NMR spectra are presented. The correlation cross-peaks from the characteristic H_5 signals assigned in the 1D-^{1}H spectrum of the mixture (red) are boxed.

The characterization of the transxyloxylation product pointed out that the xylotriose formed presented β(1-4) glycosidic linkages, revealing the enzyme regioselectivity by comparison with the NMR spectra of commercial xylotriose.

To confirm the BxTW1regioselectivity in the transxyloxylation, a reaction with *p*NPX as donor and an excess of xylose as acceptor was performed. The obtention of β(1-4)-xylobiose, identified by ^{1}H and ^{1}H-^{13}C HSQC NMR experiments, as unique transglycosylation product confirmed the regioselectivity of the enzyme.

6.2.2.2. Recombinant BxTW1

BxTW1 displayed activity against xylobiose and a high regioselectivity in catalyzing transxyloxylation reactions. However, the small amount of protein secreted by *Talaromyces amestolkiae* limited the potential application of the enzyme in the obtention of xylose derivatives through the transxyloxylation activity. The high production levels and purification yields when expressing the enzyme in *P. pastoris* permitted the obtention of a recombinant BxTW1 (rBxTW1). The improved availability of the enzyme in the recombinant form allowed to explore the scope of this enzyme as a transglycosylating catalyst for other type of acceptors. The transxyloxylation reaction by rBxTW1 in the presence of xylobiose and 2-(6-hydroxynaphthyl) (DHN) was performed. Two compounds were obtained and submitted to HPLC analysis, which suggested a major product of DHN-xylose and a minor product corresponding to a second incorporation of xylose into the DHN-xylose derivative[23].

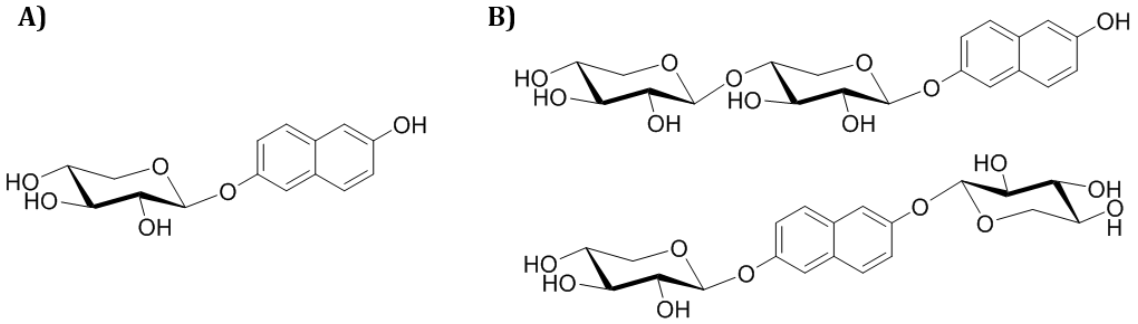


Figure 7. A) Structure of the DHN-xylose derivative. B) The two possible structures of the product result of the incorporation of a second xylose unit to the DHN-xylose.

To characterize both products, NMR experiments were performed. The ^{1}H-^{13}C HSQC of the DHN-xylose confirmed the structure (figure 7A). In the case of the minor product, the complexity of the 1D-spectum excluded the double xylosylated symmetric compound, and the ^{1}H-^{13}C HSQC demonstrated the formation of xylobiose due to the incorporation of a second xylose residue to the DHN-xylose. Furthermore, the spectrum unequivocally revealed a β(1-4) linkage, pointing out again the regioselectivity of rBxTW.

The obtained DHN-xylose is known as a selective antiproliferative compound,[26,27] acting as a decoy of the starting anchoring point for the biosynthesis of

glycosaminoglycans. The enzymatic synthesis carried out by the rBxTW1enzyme permitted the obtention of this compound avoiding the multiple protection and deprotection steps and the use of solvents employed in the traditional organic synthesis.

However, despite the advantages that these enzymatic synthesis presents in the obtention of the DHN-xylose, the use of xylobiose as donor is expensive. As stated above, xylans constitute the second most abundant type of polysaccharides in nature, formed by a backbone of β(1-4)-linked xyloses and side chains. Due its abundance and low cost, the obtention of precursors to be used as transxylosylation donor is of high relevance.

In this context, the use of cheap xylan as raw material, extracted from wood, permitted to obtain xylooligosaccharides by treatment with an endoxylanase. An enzymatic cascade strategy can be developed when rBxTW1 is also presented in the reaction mixture, the released shorter oligosaccharides could be processed by rBxTW1 and, in the presence of DHN, the DHN-xylose derivative can be obtained. DOSY-NMR experiments were key to select the beechwood xylan over the birchwood one, based on the size and the absence of natural acetylation, in order to optimize the obtention of DHN-xylose, whose production was also characterized by NMR.

6.2.2.3. Mutant BxTW1

As previously mentioned, an approach to enhance the transxyloxylation activity of a GH by blocking the hydrolysis capability consists on mutating the acid/base amino acid. The resulting mutants are called thioglycoligases.

The group of Prof. María Jesús Martínez Hernández proposed that, given that some glycosidases of the GH3 family (to which BxTW1 enzyme belongs to) are active at low pH, presumably due to the presence of an additional acid residue in the active site, this condition could allow the transxyloxylation process using many potential acceptors.

The mutation of BxTW1 by replacing the acid/base catalytic Glu^{495} by Ala (rBxTW1-E495A) resulted to be a promiscuous mutant, synthesizing a wide variety of

xylosides, not only S-xylosides but also O-, N- Se-xylosides and esters and phosphoesters derivatives. The donor used was *p*NPX, due to the necessity of a good leaving group.

A panel of potential nucleophilic acceptors with different chemical functionalities were tested: carboxylic acids, phenolic compounds, bifunctional acid and phenolic (N and N-OH nucleophiles), and P, S and Se derived compounds. In selected cases where the transglycosylation reaction occurs, the products were characterized by NMR. The ^{1}H-^{13}C HSQC and HMBC experiments were employed to confirm the formation of the xylose derivative and the position where the acceptors were glycosylated. ^{1}H-^{13}C HSQC correlates the chemical shift of two types of nuclei (in this case ^{1}H and ^{13}C) that are directly bonded, while HMBC points out the correlation through two or more chemical bonds (figure 8). The NMR spectra revealed the chemical structure of the products, presenting all of them a β linkage. The different xylose derivatives are described below, grouped according to the chemistry nature of the acceptor.

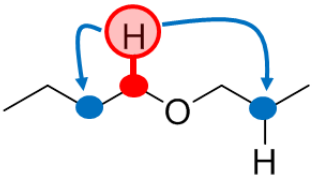


Figure 8. ^{1}H-^{13}C HSQC correlation (red) and HMBC correlation (blue).

6.2.2.3.1. Carboxylic acid

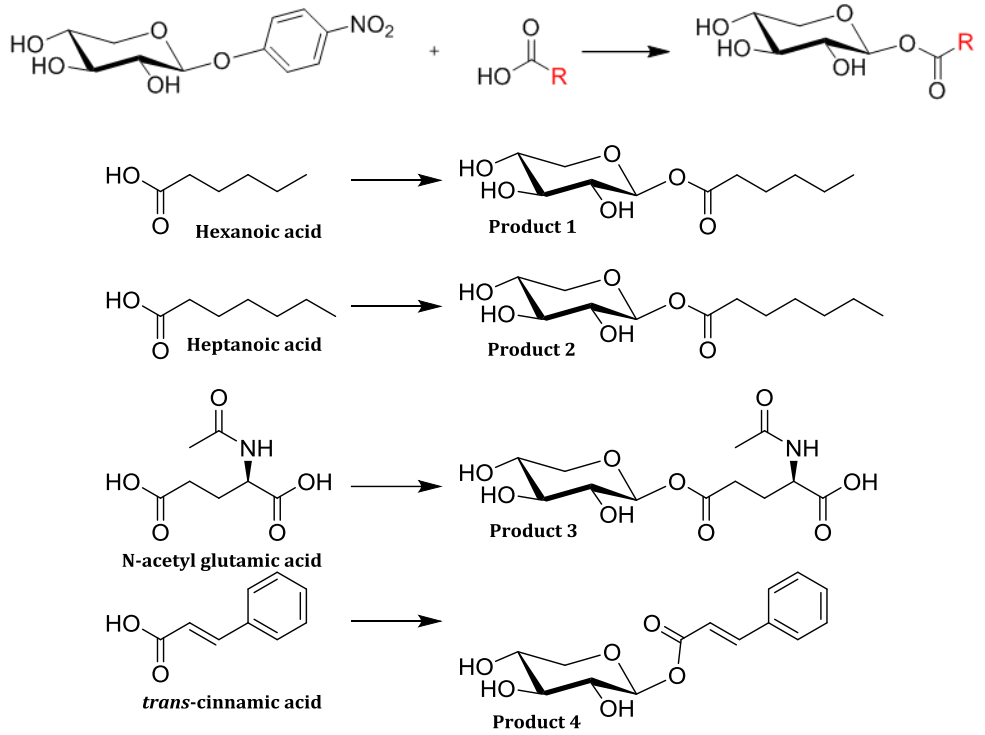


Figure 9. Top: Reaction scheme of *p*NPX and the carboxylic acid acceptor. **Bottom:** Products obtained by transxyloxylation using different acceptors.

The structural elucidation and characterization of the xylose derivatives were performed by NMR (see experimental part for the assignment tables). For instance, in the case of the N-acetyl glutamic acid, there are two carboxylic acids. In the HMBC spectra (figure 10) the correlation between the anomeric position of xylose residue (1) and the carbon of the carboxyl group (numbered C5') of the N-acetyl glutamic acid residue was observed, pointing out the formation of the derivative through this linkage. Moreover, the carbon at the carboxyl group 6' had no correlation with the xylose unit.

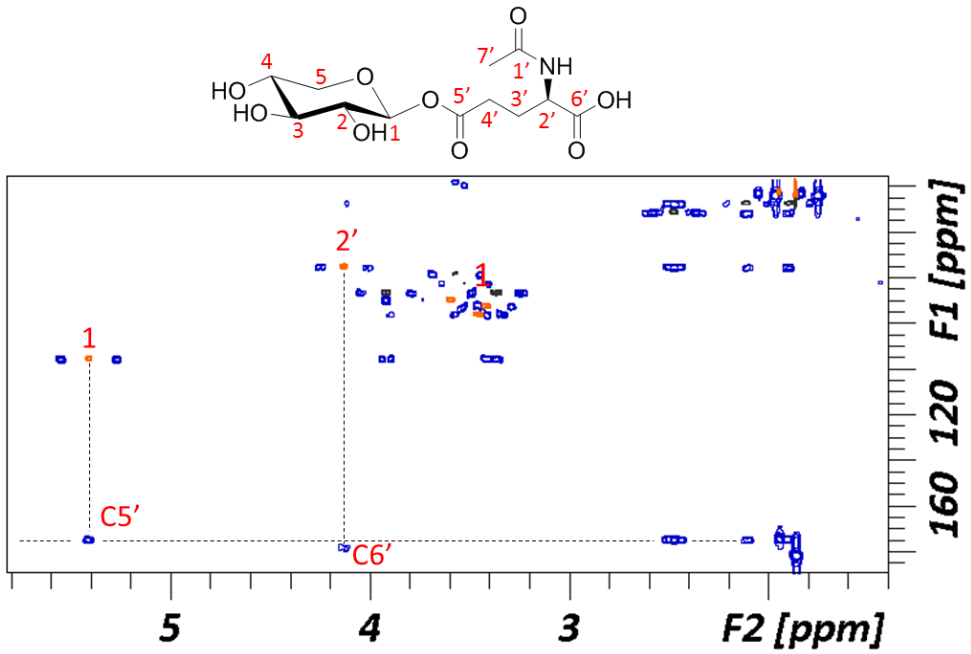


Figure 10. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the N-acetyl glutamic xylose (product **3**) derivative.

6.2.2.3.2. Phenolic alcohols

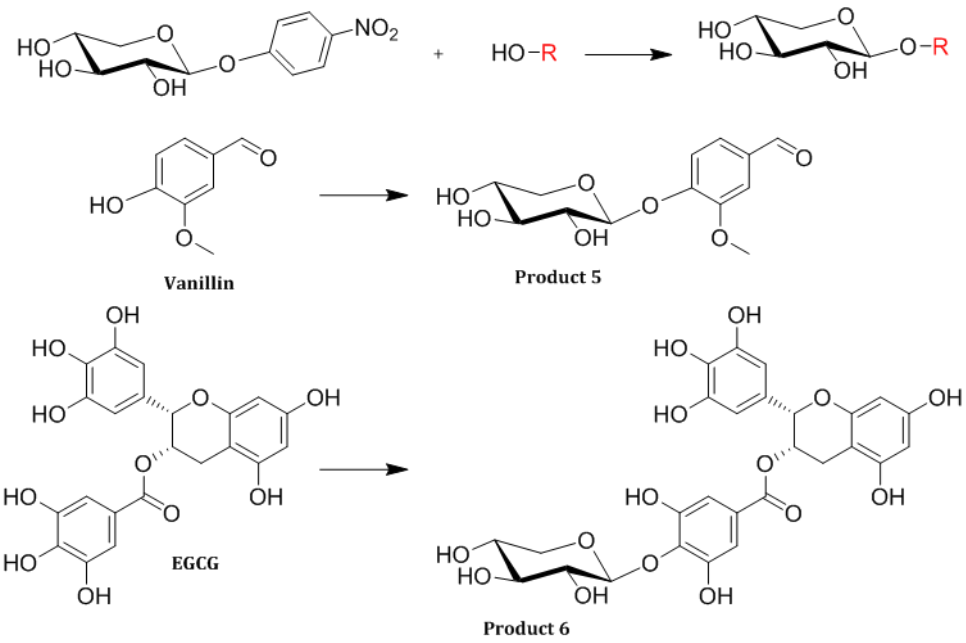


Figure 11. Top: Reaction scheme of *p*NPX and the alcohol acceptor. Products obtained by transxyloxylation using different acceptors.

The same protocol was used to characterize the alcohol derivatives. In the case of Epigallocatechin gallate (EGCG), a natural compound found in green tea with growing interest as neutracetical,[28] there are multiple -OH susceptible of

glycosylation. In order to improve its bioavailability, this catechin has been previously subject of enzymatic transglycosylations with cyclodextrin glucanotransferase, obtaining the corresponding α-glucosides either in position 3'' or 7' of the catechin rings.[29] Interestingly, in the present case, the NMR analysis (figure 12) revealed the xylose derivative formation through a β-linkage to the gallate ring, determined by the correlation between the anomeric position of xylose (1) and the carbon 4''' of the gallate moiety, widening the diversity of accessible glycosides of this interesting compound.

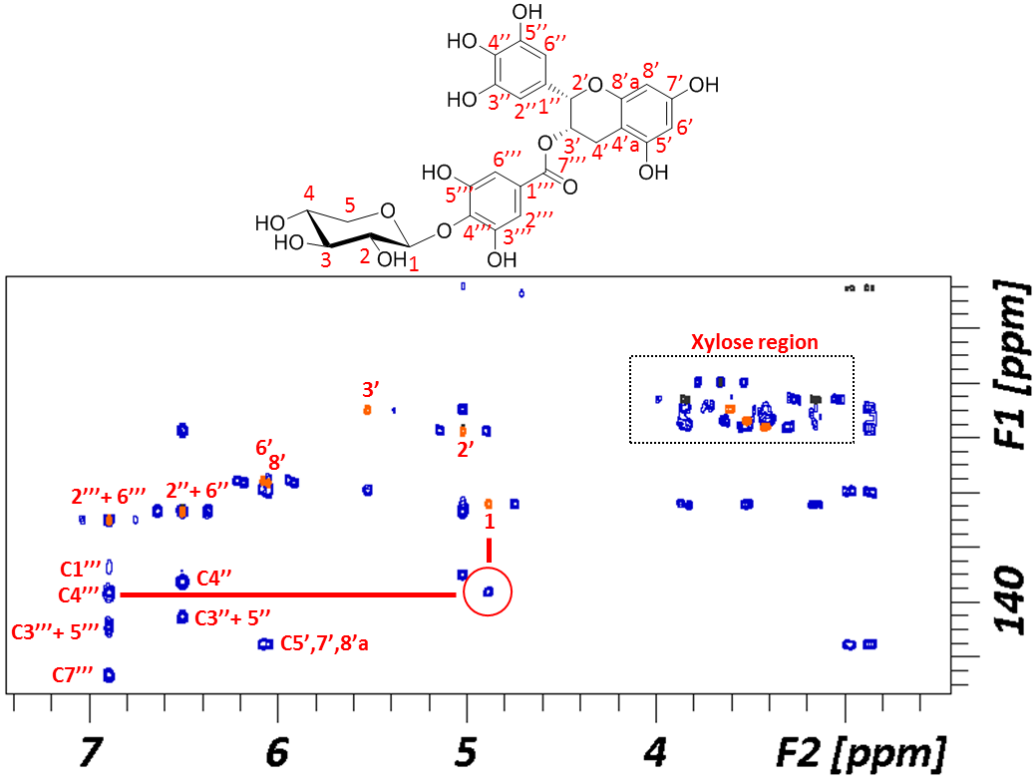


Figure 12. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the EGCG xylose derivative.

6.2.2.3.3. Carboxylic acid and alcohol

In this case, the acceptors studied, coumaric and ferulic acids, contain both alcohol and carboxylic acid groups. Short reaction times and initial low pH favored the formation of the sugar ester, while longer reaction times and pH values rendered glycosylation through the phenolic hydroxyl group. Interestingly, the double glycosylated compound could be observed (figure 13). As previously, the ^{1}H-^{13}C HSQC and HMBC experiments were employed to confirm the formation of the xylose

derivative and the position where the acceptors were glycosylated (figures 30 to 34 in appendix 1).

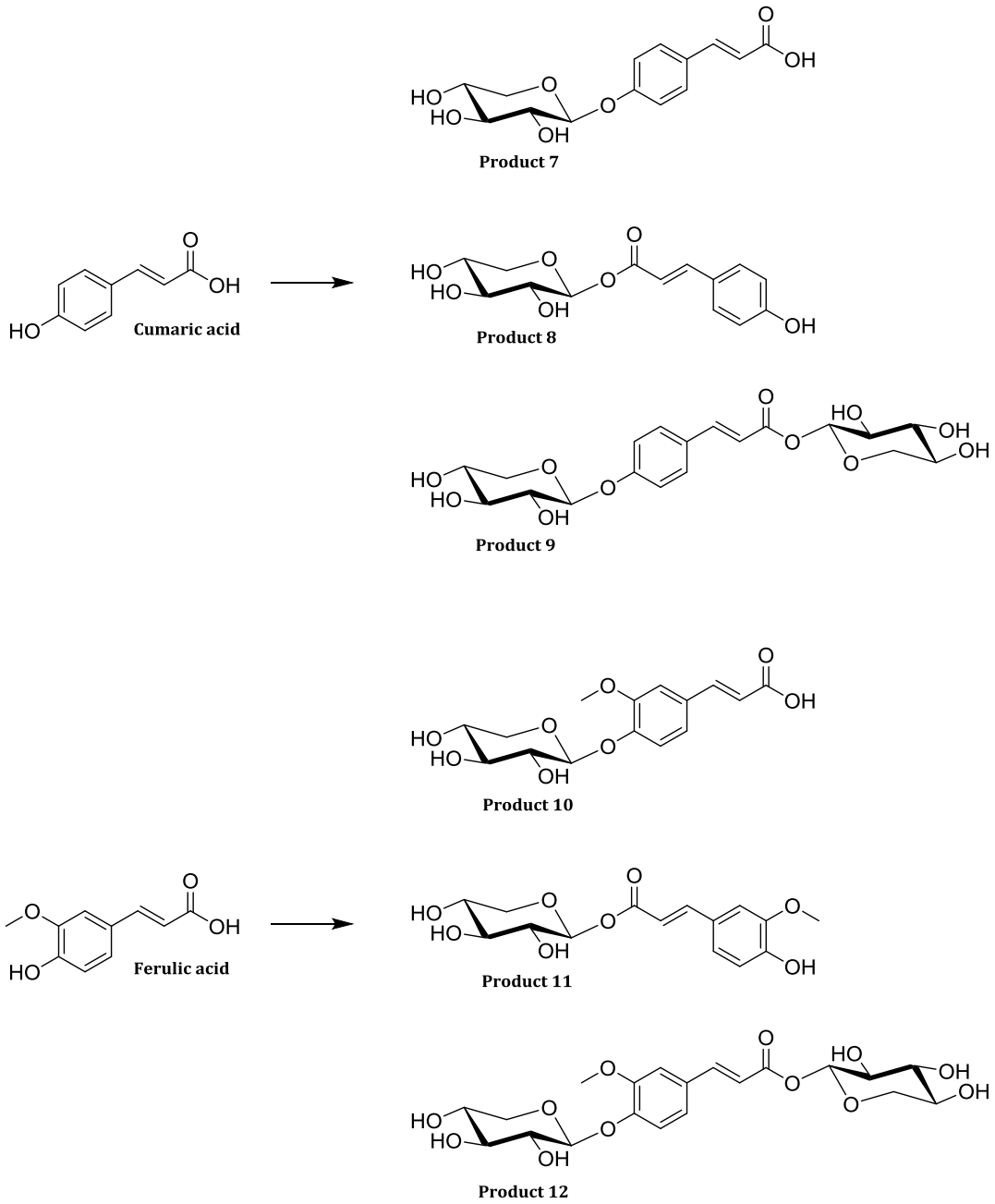


Figure 13. Products obtained by transxyloxylation using different acceptors. Three xylose derivatives were characterized for each acceptor, substitution at the carboxylic acid, at the hydroxyl or both.

6.2.2.3.4. Non-conventional C-O-N, C-O-P, C-N, C-S and C-Se glycosidic bond

The promiscuity of the mutated enzyme accepting a diversity of nucleophiles to form non-standard glycosidic bonds was proved testing a large variety of acceptors. The list comprised N-O-, N-, P-O-, S- and Se- glycosides (figure 14).

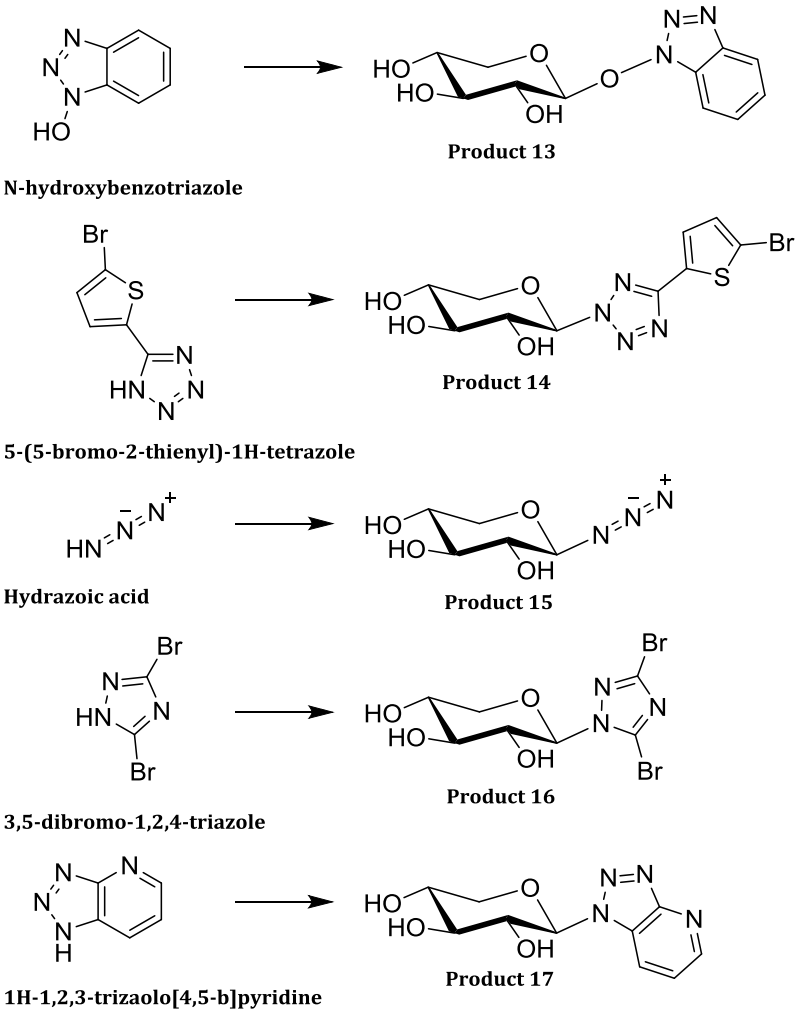


Figure 14. Products obtained by transxyloxylation using N-O and N- acceptors.

In these cases, the ^{1}H-^{13}C HSQC and HMBC approach was complicated due to the lack of ^{1}H signals in some of the compounds, hindering a direct assignment of the connectivity.

For the N-hydroxybenzotriazole derivative, the formation of the linkage was demonstrated by the correlation in the HMBC between the anomeric carbon (1) and the quaternary carbon (C6') of the HBT ring.

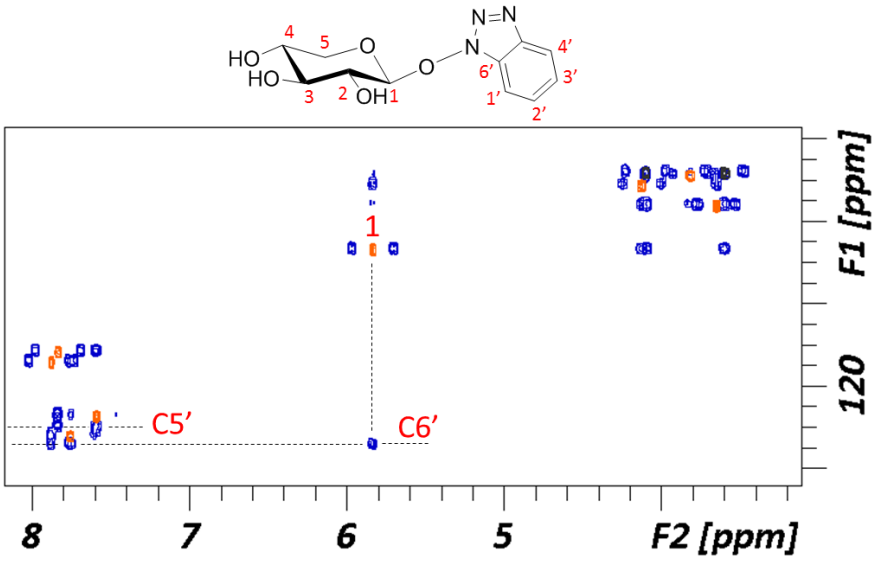


Figure 15. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the N-hydroxybenzotriazole derivative.

The 5-(5-bromo-2-thienyl)-1H-tetrazole (5BTT) derivative, product **14** (figure 16), has no ^{1}H in the tetrazole ring. The chemical shift of the anomeric position pointed out the formation of a linkage with a nitrogen of the molecule, but no discrimination between the N atoms could be done. However, the lack of a correlation between the anomeric position (1) and any position of the 5BTT molecule suggested the formation of the compound through one of the N atoms not linked to the quaternary carbon (C1').

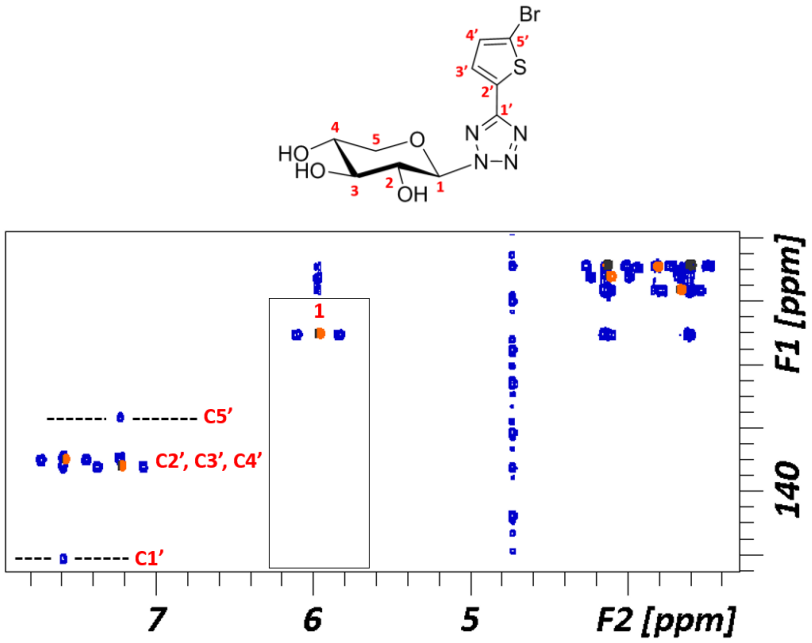


Figure 16. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the 5-(5-bromo-2-thienyl)-1H-tetrazole derivative.

For the hydrazoic acid derivative, the anomeric proton presented a coupling constant of 8.6 Hz, determining a β configuration of the linkage formed. The obtained chemical shifts were confirmed by comparison to the data previously described for the same compound[30].

The 3,5-dibromo-1,2,4-triazole (DBrT) derivative has no ^{1}H in the aglycon moiety. A correlation was observed in the HMBC between the anomeric carbon (C1) and a quaternary carbon (C3') of the DBrT ring (figure 17). However, two quaternary carbons are presented in DBrT. The acquisition of a 1D ^{13}C NMR spectrum showed the presence of two quaternary carbons with different chemical shifts. This fact, in addition to a unique correlation between C1 and C3', pointed out that the product was not symmetric. Therefore, the formation of the derivative took place through a N different from the one located between both bromine atoms.

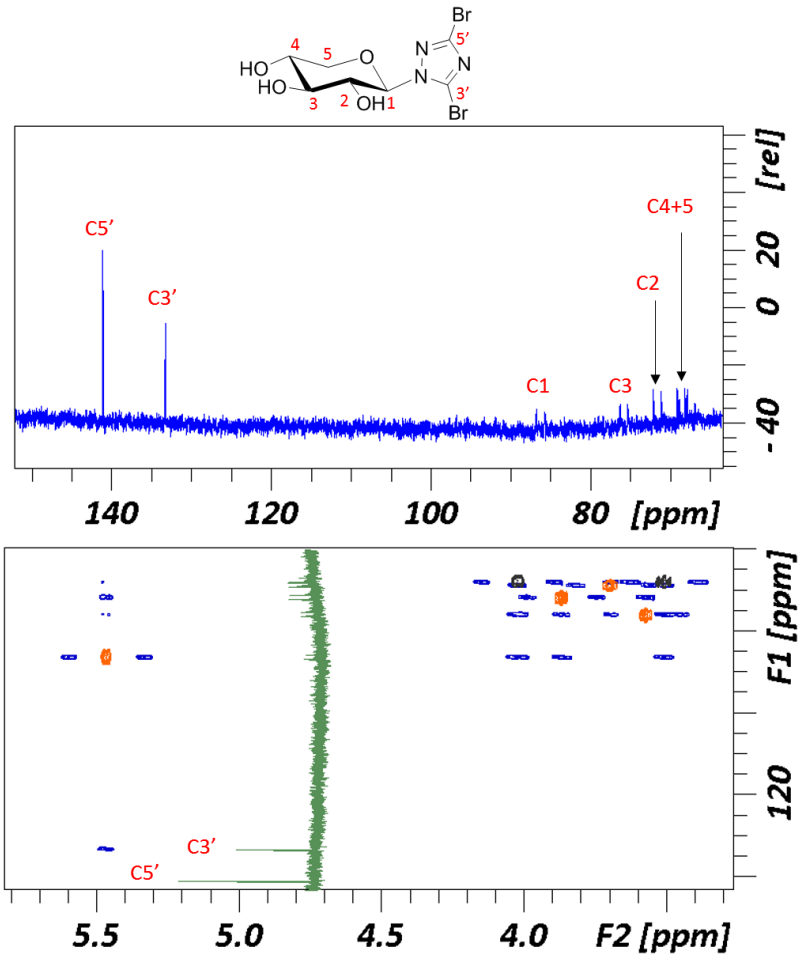


Figure 17. Top: Labelled 1D ^{13C} spectrum of the3,5-dibromo-1,2,4-triazole derivative. Duplet signals for each xylose carbons are observed due to the ^{1}H-^{13}C coupling. **Bottom:** Superimposition of ^{1}H-^{13}C HSQC (black/orange), HMBC (blue) and 1D ^{13}C (green) spectra of the3,5-dibromo-1,2,4-triazole derivative.

Finally, for the 1H-1,2,3-trizaolo[4,5-b]pyridine, a correlation between the anomeric carbon (1) and the quaternary carbon (C4') of the triazole pyridine ring was observed in the HMBC experiment, pointing out the formation of the compound through a N of the ring (figure 18). Moreover, a NOE cross peak was observed between the anomeric proton (1) and the proton 3' of the pyridine, supporting the linkage through the vicinal N to C4'.

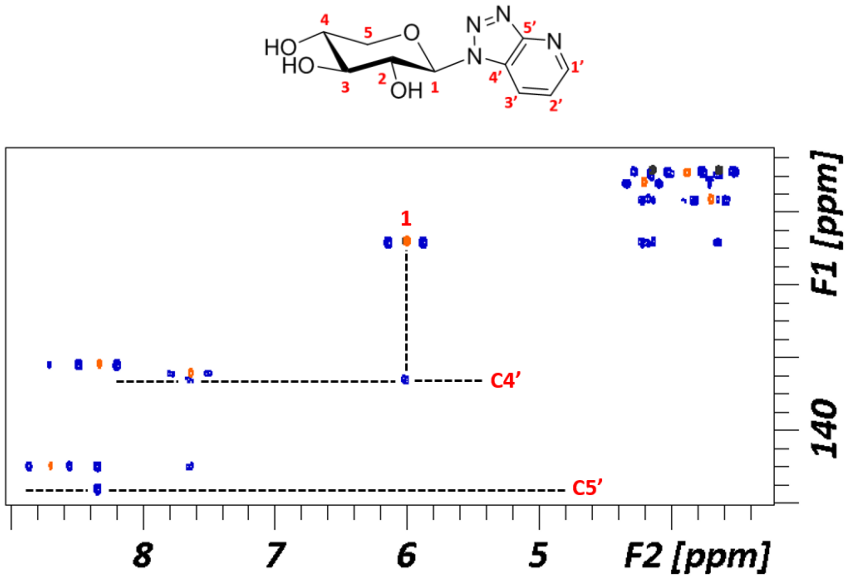


Figure 18. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the 1H-1,2,3-trizaolo[4,5-b]pyridine derivative.

The P-O-, S- and Se- glycosides were then studied (figure 19).

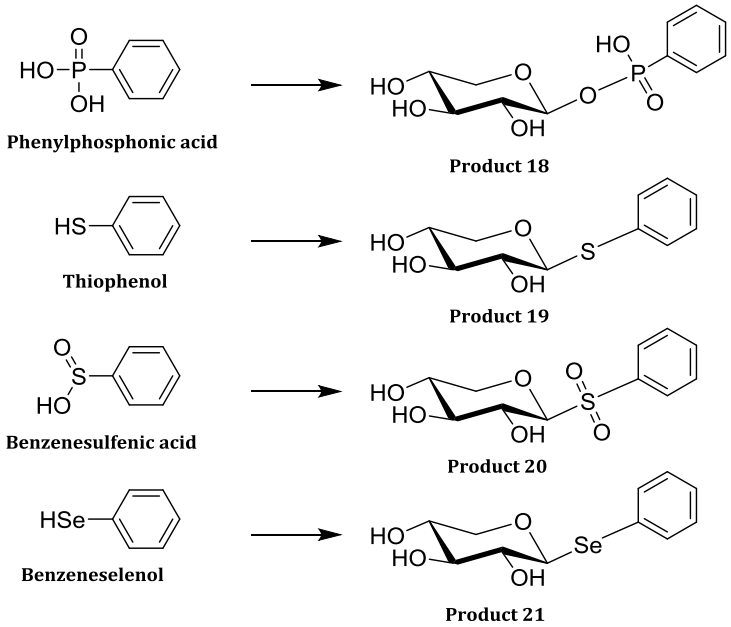


Figure 19. Products obtained by transxyloxylation using P-O-, S- and Se- acceptors.

The β linkage of the phenylphosphonic derivative (product **18**) was confirmed by the presence of two coupling constants, one between the anomeric proton and the P ($^3J_{H1-P}$= 8 Hz) and another with the H_2 ($^3J_{H1-H2}$= 8 Hz), which rendered a triplet signal instead of the regular doublet for the anomeric position.

The NMR data obtained for the thiophenol derivative (product **19**) agreed with the ones previously reported, supporting the characterization of the product[31].

In the case of the benzenesulfenic acid, the possibility of obtaining a benzenesulfinate derivative was discarded due to the chemical shift values of the anomeric position. The data agreed more to a C-S linkage instead of C-O, suggesting the formation of the sulfone (product **20**). Moreover, the NOESY experiment showed a cross-peak between the H2'/H6' aromatic protons and H1 and H2 of the xylose residue, not possible for a benzenesulfinate derivative.

Finally, the benzeneselenol derivative (product **21**) was also characterized (figure 20). The correlation between the anomeric carbon (C1) and the carbon 1 (C1') of the aromatic ring was observed in the HMBC experiment, pointing out the formation of the derivative through the linkage C1-Se- C1' linkage.

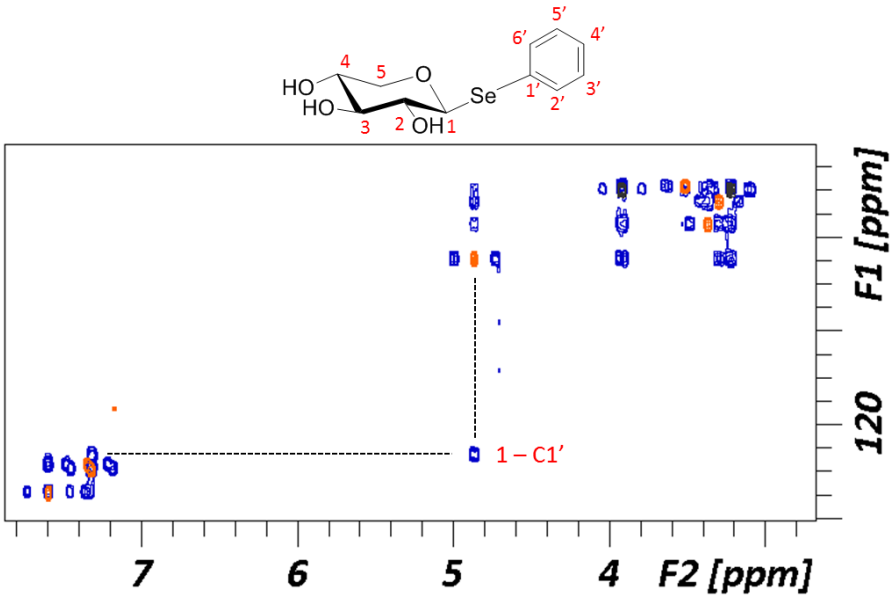


Figure 20. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the benzeneselenol derivative **20**.

6.2.3. Experimental

The native, recombinant and mutant enzymes and the isolated and purified transxyloxylation products were provided by the group of María Jesús Martinez (Biological Research Center in Madrid, CIB-CSIC). The donor xyloglycosides and the panel of acceptors were from commercial sources.

The structure and numbering of the derivatives can be found in the appendix 2.

6.2.3.1. Native BxTW1

Hydrolytic mechanism: NMR experiments were acquired at 298 K using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. The enzyme (4 μg/ml) sample was prepared in deuterated formate buffer (D_2O, formate 50 mM, pH 3) to a final concentration of 15 mM *p*NP-xylose. ^{1}H 1D NMR spectra were carried out using the corresponding sequences included in the TOPSPIN Bruker software.

Substrate specificity: NMR experiments were acquired at 298 K using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. The enzyme (26 μg/ml) sample was prepared in deuterated formate buffer (D_2O, formate 50 mM, pH 3) to a final concentration of 20 mM xylotriose. ^{1}H 1D NMR spectra, ^{1}H-^{13}C HSQC and ^{1}H-^{1}H TOCSY experiments were carried out. The corresponding pulse sequences included in the TOPSPIN Bruker software were employed.

Transxyloxylation activity: For the characterization of the transxyloxylation products, NMR experiments were acquired at 298 K using a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe. ^{1}H 1D NMR spectra, ^{1}H-^{13}C HSQC, DOSY, and DOSY-TOCSY experiments were carried out. The corresponding pulse sequences included in the TOPSPIN Bruker software were employed.

Proton	^{1}H (δ ppm)	^{13}C (δ ppm)
	α xylose	
H_1	5.19	92.22
H_2	3.51	71.49
H_3	3.66	72.86
H_4	3.61	69.32
H_5	3.68	61.03

Proton	^{1}H (δ ppm)	^{13}C (δ ppm)
	β xylose	
H_1	4.57	96.66
H_2	3.22	74.07
H_3	3.42	75.86
H_4	3.61	69.30
H_5	3.32	65.24
	3.92	

	α xylobiose	
H_1	5.17	92.11
H_2	3.54	71.51
H_3	3.74	71.08
H_4	3.76	76.63
H_5	3.74	59
	3.81	
$H_{1'}$	4.45	101.98
$H_{2'}$	3.25	72.92
$H_{3'}$	3.42	75.74
$H_{4'}$	3.62	69.31
$H_{5'}$	3.30	65.33
	3.95	

	β xylobiose	
H_1	4.58	96.62
H_2	3.24	74.09
H_3	3.54	74.08
H_4	3.76	76.63
H_5	3.37	63.14
	4.04	
$H_{1'}$	4.45	101.98
$H_{2'}$	3.25	72.92
$H_{3'}$	3.42	75.74
$H_{4'}$	3.62	69.31
$H_{5'}$	3.30	65.33
	3.95	

	α xylotriose	
H_1	5.17	92.13
H_2	3.54	71.48
H_3	3.75	71.06
H_4	3.74	76.68
H_5	3.74	58.95
	3.81	
$H_{1'}$	4.47	101.79
$H_{2'}$	3.29	72.83
$H_{3'}$	3.54	73.87
$H_{4'}$	3.77	76.54
$H_{5'}$	3.37	63.11
	4.09	
$H_{1''}$	4.44	101.98
$H_{2''}$	3.24	72.95
$H_{3''}$	3.42	75.75
$H_{4''}$	3.61	69.35
$H_{5''}$	3.30	65.32
	3.96	

	β xylotriose	
H_1	4.57	96.64
H_2	3.24	74.09
H_3	3.54	73.87
H_4	3.74	76.68
H_5	3.37	63.11
	4.04	
$H_{1'}$	4.47	101.79
$H_{2'}$	3.29	72.83
$H_{3'}$	3.54	73.87
$H_{4'}$	3.77	76.54
$H_{5'}$	3.37	63.11
	4.09	
$H_{1''}$	4.44	101.98
$H_{2''}$	3.24	72.95
$H_{3''}$	3.42	75.75
$H_{4''}$	3.61	69.35
$H_{5''}$	3.30	65.32
	3.96	

Table 1. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of α and β xylose, xylobiose and xylotriose saccharides.

Xylose: ^{1}H NMR (600 MHz, D_2O) δ 5.19 (d, *J* = 3.8 Hz, 1H), 4.57 (d, *J* = 7.8 Hz, 1H), 3.93 (dd, *J* = 11.8, 5.4 Hz, 1H), 3.72 – 3.57 (m, 5H), 3.52 (dd, *J* = 9.9, 3.7 Hz, 1H), 3.43 (t, *J* = 9.4 Hz, 1H), 3.32 (t, *J* = 11.1 Hz, 1H), 3.22 (t, *J* = 8.8 Hz, 1H).

Xylobiose: ^{1}H NMR (600 MHz, D_2O) δ 5.17 (d, *J* = 3.6 Hz, 1H), 4.58 (d, *J* = 7.8 Hz, 1H), 4.45 (d, *J* = 7.8 Hz, 2H), 4.05 (dd, *J* = 11.9, 5.3 Hz, 1H), 3.97 (dd, *J* = 11.7, 5.5 Hz, 2H),

3.84 – 3.69 (m, 5H), 3.62 (td, *J* = 9.9, 5.7 Hz, 2H), 3.54 (t, *J* = 9.1 Hz, 2H), 3.46 – 3.21 (m, 8H).

Xylotriose: ^{1}H NMR (600 MHz, D_2O) δ 5.17 (d, *J* = 3.7 Hz, 1H), 4.58 (d, *J* = 7.8 Hz, 1H), 4.46 (dd, *J* = 10.7, 7.8 Hz, 4H), 4.10 (dd, *J* = 11.9, 5.3 Hz, 2H), 4.05 (dd, *J* = 11.8, 5.4 Hz, 1H), 3.96 (dd, *J* = 11.7, 5.5 Hz, 2H), 3.86 – 3.70 (m, 7H), 3.62 (td, *J* = 9.8, 5.5 Hz, 2H), 3.55 (td, *J* = 9.1, 4.3 Hz, 4H), 3.42 (t, *J* = 9.2 Hz, 2H), 3.37 (t, *J* = 10.7 Hz, 3H), 3.27 (dtd, *J* = 26.1, 9.9, 8.6, 4.9 Hz, 7H).

6.2.3.2. Recombinant BxTW1: DHN transxyloxylation products characterization

NMR experiments were acquired at 308 K using a Bruker AVANCE 500 MHz spectrometer. The samples were prepared in 500 µL of deuterated water (D_2O). ^{1}H 1D NMR, ^{1}H-^{13}C HSQC and ^{1}H-^{1}H TOCSY spectra were carried out using the corresponding sequences included in the TOPSPIN Bruker software.

	^{1}H (δ ppm)	^{13}C (δ ppm)
	DHN-xylose	
H_1	5.22	101.35
H_2	3.65	73.08
H_3	3.65	75.77
H_4	3.79	69.32
H_5	3.57	65.45
	4.10	
$H_{1'}$	7.53	111.82
$H_{3'}$	7.34	119.53
$H_{4'}$	7.84	129.31
$H_{5'}$	7.32	109.65
$H_{7'}$	7.25	119.17
$H_{8'}$	7.82	128.62

	^{1}H (δ ppm)	^{13}C (δ ppm)
	DHN-xylobiose	
H_1	5.23	101.24
H_2	3.69	72.92
H_3	3.76	73.83
H_4	3.94	76.45
H_5	3.64	63.23
	4.25	
$H_{1'}$	4.54	102.13
$H_{2'}$	3.34	72.98
$H_{3'}$	3.49	75.86
$H_{4'}$	3.69	69.38
$H_{5'}$	3.37	65.38
	4.05	
$H_{1''}$	7.53	111.75
$H_{3''}$	7.34	119.53
$H_{4''}$	7.84	129.15
$H_{5''}$	7.31	109.60
$H_{7''}$	7.24	119.29
$H_{8''}$	7.82	128.61

Table 2. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of the transxyloxylation products DHN-xylose and DHN-xylobiose.

DHN-xylose: ^{1}H NMR (500 MHz, D_2O) δ 7.83 (dd, *J* = 12.7, 9.0 Hz, 2H), 7.52 (s, 1H), 7.37 – 7.30 (m, 2H), 7.24 (d, *J* = 9.0 Hz, 1H), 5.22 (d, *J* = 6.3 Hz, 1H), 4.09 (dd, *J* = 11.6, 5.2 Hz, 1H), 3.79 (m, 1H), 3.64 (d, *J* = 4.1 Hz, 2H), 3.57 (d, *J* = 11.0 Hz, 1H).

DHN-xylobiose: ^{1}H NMR (500 MHz, D_2O) δ 7.82 (dd, *J* = 12.1, 8.9 Hz, 2H), 7.52 (s, 1H), 7.37 – 7.29 (m, 2H), 7.24 (d, *J* = 8.9, 1H), 5.23 (d, *J* = 7.5 Hz, 1H), 4.54 (d, *J* = 7.8 Hz, 1H), 4.23 (dd, *J* = 11.9, 5.2 Hz, 1H), 4.04 (dd, *J* = 11.7, 5.5 Hz, 1H), 3.94 (td, *J* = 9.5, 5.1 Hz, 1H), 3.76 (t, *J* = 9.0 Hz, 1H), 3.73 – 3.59 (m, 3H), 3.49 (t, *J* = 9.1 Hz, 1H), 3.41 – 3.30 (m, 2H).

6.2.3.3. Mutant BxTW1: transxyloxylation products characterization.

NMR experiments of the purified compounds were acquired at 298 K using a Bruker AVANCE 500 MHz spectrometer and a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe (magnetic field indicated in the characterization section). Samples were prepared in 500 µL of deuterated water (D_2O). 1D ^{1}H NMR spectra, ^{1}H-^{13}C HSQC, HMBC, ^{1}H-^{1}H TOCSY and selective TOCSY experiments were carried out in order to fully assign the NMR signals of the compounds. The standard pulse sequences included in the TOPSPIN Bruker software were employed.

	^{1}H (δ ppm)	^{13}C (δ ppm)
	Product 1	
H_1	5.40	94.43
H_2	3.40	71.49
H_3	3.44	75.26
H_4	3.59	68.66
H_5	3.35	65.67
	3.90	
$H_{1'}$	---	175.48
$H_{2'}$	2.40	33.45
$H_{3'}$	1.56	23.27
$H_{4'}$	1.23	30.22
$H_{5'}$	1.23	21.21
$H_{6'}$	0.79	14.27

	^{1}H (δ ppm)	^{13}C (δ ppm)
	Product 2	
H_1	5.40	94.87
H_2	3.40	72.16
H_3	3.47	75.70
H_4	3.59	69.26
H_5	3.38	65.86
	3.92	
$H_{1'}$	---	175.49
$H_{2'}$	2.41	33.59
$H_{3'}$	1.56	24.10
$H_{4'}$	1.26	28.32
$H_{5'}$	1.21	31.42
$H_{6'}$	1.23	21.29
$H_{7'}$	0.78	11.57

	^{1}H (δ ppm)	^{13}C (δ ppm)
	Product 11	
H_1	5.51	94.77
H_2	3.50	71.90
H_3	3.50	75.35
H_4	3.63	69.01
H_5	3.41	65.80
	3.95	
$H_{1'}$	---	126.69
$H_{2'}$	7.15	111.36
$H_{3'}$	---	147.60
$H_{4'}$	---	147.60
$H_{5'}$	6.84	115.53
$H_{6'}$	7.07	123.76
$H_{7'}$	7.66	147.97
$H_{8'}$	6.35	113.18
$H_{9'}$	---	167.71
$H_{10'}$	3.80	55.82

	Product 12	
H_1	5.05	100.65
H_2	3.56	72.65
H_3	3.50	75.33
H_4	3.65	68.96
H_5	3.40	65.55
	3.95	
$H_{1'}$	---	129.38
$H_{2'}$	7.23	111.72
$H_{3'}$	---	148.97

	Product 3	
H_1	5.41	94.74
H_2	3.41	71.70
H_3	3.44	75.41
H_4	3.59	68.83
H_5	3.36	65.77
	3.90	
$H_{1'}$	---	173.92
$H_{2'}$	4.13	54.32
$H_{3'}$	1.89	26.53
	2.10	
$H_{4'}$	2.46	30.51
$H_{5'}$	---	173.92
$H_{6'}$	---	177.39
$H_{7'}$	1.94	21.61

	Product 4	
H_1	5.53	94.85
H_2	3.51	71.26
H_3	3.51	75.42
H_4	3.63	69.00
H_5	3.41	65.79
	3.95	
$H_{1'}$	---	133.81
$H_{2'}$	7.62	128.55
$H_{3'}$	7.42	129.08
$H_{4'}$	7.42	131.28
$H_{5'}$	7.42	129.08
$H_{6'}$	7.62	128.55
$H_{7'}$	---	167.14
$H_{8'}$	6.56	116.16
$H_{9'}$	7.81	148.03

	Product 5	
H_1	5.15	100.21
H_2	3.59	72.67
H_3	3.51	75.21
H_4	3.66	69.17
H_5	3.44	65.74
	3.95	
$H_{1'}$	---	150.70
$H_{2'}$	---	149.26
$H_{3'}$	7.50	111.80
$H_{4'}$	---	131.31
$H_{5'}$	7.52	126.60
$H_{6'}$	7.23	115.12
$H_{7'}$	9.72	194.37
$H_{8'}$	3.86	56.23

	Product 6	
H_1	4.87	103.8
H_2	3.51	73.33
H_3	3.41	75.42
H_4	3.60	69.20
$H_{4'}$	---	147.58
$H_{5'}$	7.09	115.84
$H_{6'}$	7.16	123.13
$H_{7'}$	7.71	147.46
$H_{8'}$	6.43	114.86
$H_{9'}$	---	167.42
$H_{10'}$	3.82	55.94
$H_{1''}$	5.53	94.81
$H_{2''}$	3.52	71.97
$H_{3''}$	3.50	75.33
$H_{4''}$	3.65	69.03
$H_{5''}$	3.40	65.35
	3.95	

	Product 13	
H_1	5.82	86.80
H_2	4.10	71.19
H_3	3.64	76.08
H_4	3.81	68.66
H_5	3.59	68.03
	4.09	
$H_{1'}$	7.87	114.40
$H_{2'}$	7.75	132.20
$H_{3'}$	7.58	127.39
$H_{4'}$	7.83	111.53
$H_{5'}$	---	129.88
$H_{6'}$	---	134.55

	Product 14	
H_1	5.95	89.83
H_2	4.09	71.80
H_3	3.64	75.91
H_4	3.80	68.70
H_5	3.59	68.01
	4.11	
$H_{1'}$	---	160.71
$H_{2'}$	---	129.95
$H_{3'}$	7.58	129.47
$H_{4'}$	7.22	131.53
$H_{5'}$	---	116.02

	Product 15	
H_1	4.60	90.90
H_2	3.17	72.74
H_3	3.38	75.80
H_4	3.56	68.96
H_5	3.31	66.90
	3.95	

	Product 16	
H_1	5.46	86.49
H_2	3.86	72.16
H_3	3.57	76.29
H_4	3.69	68.96
H_5	3.51	67.85

H_5	3.15	65.54
	3.85	
$H_{2'}$	5.01	77.18
$H_{3'}$	5.52	69.04
$H_{4'}$	2.87	25.13
	2.98	
$H_{4a'}$	---	99.41
$H_{5'}$	---	154.95
$H_{6'}$	6.07	95.30
$H_{7'}$	---	154.95
$H_{8'}$	6.04	96.07
$H_{8a'}$	---	154.95
$H_{1''}$	---	129.5
$H_{2''}$	6.50	106.46
$H_{3''}$	---	145.14
$H_{4''}$	---	132.15
$H_{5''}$	---	145.14
$H_{6''}$	6.50	106.46
$H_{1'''}$	---	127.06
$H_{2'''}$	6.89	109.70
$H_{3'''}$	---	148.53
$H_{4'''}$	---	135.99
$H_{5'''}$	---	148.53
$H_{6'''}$	6.89	109.70
$H_{7'''}$	---	165.80

	Product 7	
H_1	5.08	100.27
H_2	3.51	72.74
H_3	3.50	75.40
H_4	3.66	68.99
H_5	3.44	65.18
	3.97	
$H_{2'}$	7.58	129.98
$H_{3'}$	7.08	116.75
$H_{5'}$	7.08	116.75
$H_{6'}$	7.58	129.98
$H_{7'}$	7.58	144.25
$H_{8'}$	6.41	117.65

	Product 8	
H_1	5.51	94.71
H_2	3.50	71.85
H_3	3.51	75.39
H_4	3.63	69.01
H_5	3.40	65.83
	3.95	
$H_{2'}$	7.53	130.76
$H_{3'}$	6.88	115.84
$H_{5'}$	6.88	115.84
$H_{6'}$	7.53	130.76
$H_{7'}$	7.76	147.95
$H_{8'}$	6.38	113.05

	Product 9	
H_1	5.08	100.17
	4.02	
$H_{2'}$	---	133.28
$H_{4'}$	---	141.06

	Product 17	
H_1	6.00	87.66
H_2	4.20	71.43
H_3	3.69	76.13
H_4	3.88	68.76
H_5	3.62	68.02
	4.13	
$H_{1'}$	8.70	149.56
$H_{2'}$	7.63	123.95
$H_{3'}$	8.33	121.47
$H_{4'}$	---	125.68
$H_{5'}$	---	155.69

	Product 18	
H_1	4.74	97.59
H_2	3.22	73.47
H_3	3.32	75.19
H_4	3.50	69.10
H_5	3.17	65.55
	3.76	
$H_{2'}$	7.68	130.53
$H_{3'}$	7.41	128.23
$H_{4'}$	7.48	130.88
$H_{5'}$	7.41	128.23
$H_{6'}$	7.68	130.53

	Product 19	
H_1	4.66	88.06
H_2	3.26	71.67
H_3	3.40	77.02
H_4	3.53	69.05
H_5	3.24	68.76
	3.90	
$H_{2'}$	7.49	132.08
$H_{3'}$	7.34	128.91
$H_{4'}$	7.34	128.91
$H_{5'}$	7.34	128.91
$H_{6'}$	7.49	132.08

	Product 20	
H_1	4.57	91.29
H_2	3.60	69.28
H_3	3.42	76.85
H_4	3.48	68.31
H_5	3.20	69.31
	3.90	
$H_{2'}$	7.86	129.10
$H_{3'}$	7.62	129.56
$H_{4'}$	7.75	135.39
$H_{5'}$	7.62	129.56
$H_{6'}$	7.86	129.10

H_2	3.52	72.70
H_3	3.51	75.47
H_4	3.64	69.03
H_5	3.44	65.27
	3.96	
$H_{2'}$	7.60	130.47
$H_{3'}$	7.09	116.75
$H_{5'}$	7.09	116.75
$H_{6'}$	7.60	130.47
$H_{7'}$	7.77	147.39
$H_{8'}$	6.44	114.73
$H_{1''}$	5.53	94.81
$H_{2''}$	3.51	71.85
$H_{3''}$	3.51	75.47
$H_{4''}$	3.64	69.03
$H_{5''}$	3.44	65.82
	3.97	

	Product 10	
H_1	5.04	100.74
H_2	3.54	72.71
H_3	3.48	75.41
H_4	3.65	69.07
H_5	3.39	65.35
	3.95	
$H_{1'}$	---	129.76
$H_{2'}$	7.21	111.60
$H_{3'}$	---	148.97
$H_{4'}$	---	147.19
$H_{5'}$	7.07	115.90
$H_{6'}$	7.13	122.57
$H_{7'}$	7.55	145.57
$H_{8'}$	6.35	116.57
$H_{9'}$	---	171.28
$H_{10'}$	3.82	55.89

	Product 21	
H_1	4.86	84.62
H_2	3.29	72.25
H_3	3.37	77.02
H_4	3.51	68.93
H_5	3.22	69.86
	3.92	
$H_{1'}$	---	126.54
$H_{2'}$	7.59	134.52
$H_{3'}$	7.32	129.40
$H_{4'}$	7.35	128.65
$H_{5'}$	7.32	129.40
$H_{6'}$	7.59	134.52

Table 3. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of the transxyloxylation products **1** to **21** (see appendix 2 figure 35 for molecule structures and atom numbering).

Product 1: ^{1}H NMR (600 MHz, D_2O) δ 5.40 (d, *J* = 7.9 Hz, 1H), 3.90 (dd, *J* = 11.6, 5.4 Hz, 1H), 3.59 (ddd, *J* = 10.4, 8.8, 5.4 Hz, 1H), 3.48 – 3.33 (m, 3H), 2.40 (t, *J* = 7.4 Hz, 2H), 1.56 (p, *J* = 7.1 Hz, 2H), 1.23 (dq, *J* = 7.2, 3.8, 3.3 Hz, 4H), 0.82 – 0.76 (m, 3H).

Product 2: ^{1}H NMR (600 MHz, D_2O) δ 5.40 (d, *J* = 7.9 Hz, 1H), 3.92 (dd, *J* = 11.6, 5.4 Hz, 1H), 3.59 (td, *J* = 9.6, 5.4 Hz, 1H), 3.49 – 3.33 (m, 3H), 2.41 (t, *J* = 7.4 Hz, 2H), 1.56 (p, *J* = 7.3 Hz, 2H), 1.29 – 1.18 (m, 6H), 0.81 – 0.76 (m, 3H).

Product 3: ^{1}H NMR (600 MHz, D_2O) δ 5.41 (d, *J* = 7.7 Hz, 1H), 4.13 (dd, *J* = 8.9, 4.7 Hz, 1H), 3.9 (dd, *J* = 11.7, 5.4 Hz, 1H), 3.63 – 3.53 (m, 1H), 3.48 – 3.33 (m, 3H), 2.46 (q, *J* = 7.8 Hz, 2H), 2.14 – 2.07 (m, 1H), 1.94 (s, 3H), 1.93 – 1.87 (m, 1H).

Product 4: ^{1}H NMR (600 MHz, D_2O) δ 7.81 (d, *J* = 16.1 Hz, 1H), 7.62 (dd, *J* = 7.3, 1.9 Hz, 2H), 7.47 – 7.38 (m, 3H), 6.56 (d, *J* = 16.0 Hz, 1H), 5.57 – 5.50 (m, 1H), 3.95 (dd, *J* = 11.6, 5.4 Hz, 1H), 3.63 (ddd, *J* = 10.4, 8.3, 5.3 Hz, 1H), 3.55 – 3.47 (m, 2H), 3.41 (t, *J* = 11.0 Hz, 1H).

Product 5: ^{1}H NMR (600 MHz, D_2O) δ 9.72 (s, 1H), 7.55 – 7.48 (m, 2H), 7.23 (d, *J* = 8.3 Hz, 1H), 5.15 (d, *J* = 7.6 Hz, 1H), 3.95 (dd, *J* = 11.6, 5.4 Hz, 1H), 3.86 (s, 3H), 3.70 – 3.55 (m, 2H), 3.51 (t, *J* = 9.2 Hz, 1H), 3.47 – 3.40 (m, 1H).

Product 6: ^{1}H NMR (600 MHz, D_2O) δ 6.89 (s, 2H), 6.50 (s, 2H), 6.09 – 6.03 (m, 2H), 5.54 – 5.50 (m, 1H), 5.01 (s, 1H), 4.87 (d, *J* = 7.7 Hz, 1H), 3.85 (dd, *J* = 11.7, 5.4 Hz, 1H), 3.67 – 3.56 (m, 1H), 3.51 (dd, *J* = 9.4, 7.7 Hz, 1H), 3.41 (t, *J* = 9.2 Hz, 1H), 3.15 (t, *J* = 11.0 Hz, 1H), 2.98 (dd, *J* = 17.5, 4.5 Hz, 1H), 2.89 – 2.83 (m, 1H).

Product 7: ^{1}H NMR (500 MHz, D_2O) δ 7.60 – 7.53 (m, 3H), 7.08 (d, *J* = 8.2 Hz, 2H), 6.41 (d, *J* = 16.3 Hz, 1H), 5.08 (d, *J* = 5.3 Hz, 1H), 3.97 (dd, *J* = 11.9, 5.2 Hz, 1H), 3.66 (d, *J* = 8.7 Hz, 1H), 3.53 – 3.39 (m, 3H).

Product 8: ^{1}H NMR (500 MHz, D_2O) δ 7.76 (d, *J* = 16.0 Hz, 1H), 7.56 – 7.49 (m, 2H), 6.91 – 6.84 (m, 2H), 6.38 (d, *J* = 15.9 Hz, 1H), 5.54 – 5.47 (m, 1H), 3.95 (dd, *J* = 11.5, 5.4 Hz, 1H), 3.66 – 3.59 (m, 1H), 3.55 – 3.46 (m, 2H), 3.4 (dd, *J* = 11.7, 10.3 Hz, 1H).

Product 9: ^{1}H NMR (500 MHz, D_2O) δ 7.77 (d, *J* = 16.0 Hz, 1H), 7.64 – 7.57 (m, 2H), 7.13 – 7.02 (m, 2H), 6.44 (d, *J* = 16.0 Hz, 1H), 5.57 – 5.49 (m, 1H), 5.13 – 5.04 (m, 1H), 3.96 (ddd, *J* = 11.6, 5.4, 1.9 Hz, 2H), 3.64 (dddd, *J* = 10.3, 8.9, 7.0, 5.1 Hz, 2H), 3.55 – 3.36 (m, 6H).

Product 10: ^{1}H NMR (600 MHz, D_2O) δ 7.55 (d, *J* = 15.9 Hz, 1H), 7.21 (d, *J* = 2.0 Hz, 1H), 7.13 (dd, *J* = 8.4, 1.9 Hz, 1H), 7.07 (d, *J* = 8.4 Hz, 1H), 6.35 (d, *J* = 16.1 Hz, 1H), 5.04 (d, *J* = 7.7 Hz, 1H), 3.95 (dd, *J* = 11.6, 5.5 Hz, 1H), 3.82 (s, 3H), 3.69 – 3.62 (m, 1H), 3.58 – 3.46 (m, 2H), 3.39 (t, *J* = 11.0 Hz, 1H).

Product 11: ^{1}H NMR (600 MHz, D_2O) δ 7.66 (d, *J* = 16.0 Hz, 1H), 7.15 (d, *J* = 2.0 Hz, 1H), 7.07 (dd, *J* = 8.2, 1.9 Hz, 1H), 6.84 (d, *J* = 8.2 Hz, 1H), 6.35 (d, *J* = 15.9 Hz, 1H), 5.54 – 5.47 (m, 1H), 3.95 (dd, *J* = 11.6, 5.5 Hz, 1H), 3.80 (s, 3H), 3.67 – 3.59 (m, 1H), 3.54 – 3.46 (m, 2H), 3.41 (dd, *J* = 11.7, 10.3 Hz, 1H).

Product 12: ^{1}H NMR (600 MHz, D_2O) δ 7.71 (d, *J* = 16.0 Hz, 1H), 7.23 (s, 1H), 7.16 (d, *J* = 8.4 Hz, 1H), 7.09 (d, *J* = 8.3 Hz, 1H), 6.43 (d, *J* = 15.9 Hz, 1H), 5.53 (d, *J* = 7.0 Hz, 1H), 5.05 (d, *J* = 7.6 Hz, 1H), 3.95 (td, *J* = 12.1, 5.4 Hz, 2H), 3.82 (s, 3H), 3.65 (qd, *J* = 9.1, 8.7, 5.1 Hz, 2H), 3.58 – 3.46 (m, 4H), 3.40 (q, *J* = 11.8 Hz, 2H).

Product 13: ^{1}H NMR (600 MHz, D_2O) δ 7.85 (dd, *J* = 24.9, 8.6 Hz, 2H), 7.75 (t, *J* = 7.8 Hz, 1H), 7.58 (dd, *J* = 8.6, 7.0 Hz, 1H), 5.82 (d, *J* = 9.1 Hz, 1H), 4.15 – 4.06 (m, 2H), 3.81 (td, *J* = 9.9, 5.5 Hz, 1H), 3.68 – 3.56 (m, 2H).

Product 14: ^{1}H NMR (600 MHz, D_2O) δ 7.58 (d, *J* = 4.0 Hz, 1H), 7.22 (d, *J* = 4.0 Hz, 1H), 5.95 (d, *J* = 9.1 Hz, 1H), 4.15 – 4.06 (m, 2H), 3.80 (ddd, *J* = 10.7, 9.2, 5.5 Hz, 1H), 3.68 – 3.55 (m, 2H).

Product 15: ^{1}H NMR (600 MHz, D_2O) δ 4.60 (d, *J* = 8.7 Hz, 1H), 3.95 (dd, *J* = 11.6, 5.5 Hz, 1H), 3.56 (ddd, *J* = 10.6, 9.1, 5.5 Hz, 1H), 3.41 – 3.28 (m, 2H), 3.17 (t, *J* = 9.0 Hz, 1H).

Product 16: ^{1}H NMR (600 MHz, D_2O) δ 5.46 (d, *J* = 8.9 Hz, 1H), 4.02 (dd, *J* = 11.5, 5.5 Hz, 1H), 3.86 (t, *J* = 9.1 Hz, 1H), 3.69 (ddd, *J* = 10.6, 9.3, 5.5 Hz, 1H), 3.60 – 3.47 (m, 2H).

Product 17: ^{1}H NMR (600 MHz, D_2O) δ 8.70 (d, J = 4.6 Hz, 1H), 8.33 (d, J = 8.5 Hz, 1H), 7.63 (dd, J = 8.5, 4.5 Hz, 1H), 6.00 (d, J = 9.2 Hz, 1H), 4.20 (t, J = 9.2 Hz, 1H), 4.13 (dd, J = 11.5, 5.6 Hz, 1H), 3.88 (ddd, J = 10.8, 9.2, 5.5 Hz, 1H), 3.72 – 3.59 (m, 2H).

Product 18: ^{1}H NMR (600 MHz, D_2O) δ 7.68 (dd, *J* = 12.7, 7.8 Hz, 2H), 7.57 – 7.33 (m, 3H), 4.74 (d, *J* = 7.9 Hz, 1H), 3.82 – 3.73 (m, 1H), 3.56 – 3.44 (m, 1H), 3.32 (t, *J* = 9.1 Hz, 1H), 3.27 – 3.12 (m, 2H).

Product 19: ^{1}H NMR (600 MHz, D_2O) δ 7.49 (dd, *J* = 7.3, 2.2 Hz, 2H), 7.38 – 7.30 (m, 3H), 4.66 (d, *J* = 9.7 Hz, 1H), 3.90 (dd, *J* = 11.4, 5.4 Hz, 1H), 3.53 (td, *J* = 9.9, 5.4 Hz, 1H), 3.40 (t, *J* = 9.0 Hz, 1H), 3.25 (q, *J* = 11.3, 10.3 Hz, 2H).

Product 20: ^{1}H NMR (600 MHz, D_2O) δ 7.86 (d, *J* = 7.6 Hz, 2H), 7.75 (t, *J* = 7.5 Hz, 1H), 7.62 (t, *J* = 7.8 Hz, 2H), 4.57 (d, *J* = 9.5 Hz, 1H), 3.90 (dd, *J* = 11.4, 5.4 Hz, 1H), 3.60 (t, *J* = 9.2 Hz, 1H), 3.52 – 3.39 (m, 2H), 3.20 (t, *J* = 10.9 Hz, 1H).

Product 21: ^{1}H NMR (600 MHz, D_2O) δ 7.63 – 7.57 (m, 2H), 7.39 – 7.29 (m, 3H), 4.86 (dd, *J* = 9.6, 0.8 Hz, 1H), 3.92 (dd, *J* = 11.4, 5.3 Hz, 1H), 3.58 – 3.43 (m, 1H), 3.37 (t, *J* = 8.9 Hz, 1H), 3.29 (ddd, *J* = 9.4, 8.7, 0.7 Hz, 1H), 3.25 – 3.18 (m, 1H).

6.3. β-GLUCOSIDASE FROM *TALAROMYCES AMESTOLKIAE*

6.3.1. Introduction

As previously mentioned, plant biomass, the main renewable resource in biosphere, is composed of two polysaccharides, cellulose and hemicellulose. Being cellulose the most abundant biopolymer in nature, it is a very attractive raw material for transformation in commodity products with very important economic repercussions all over the world, as happens with bioethanol. The first step in this process is to depolymerize cellulose to its glucose monomer, which can be then fermented to produce ethanol. In this depolymerization process, different glycosidases (cellulases, glucosidases) play a key role.

β-glucosidases (BGL) catalyze the transformation of cellobiose or cellooligosaccharides into glucose. This is of significant importance for obtaining glucose monosaccharide from plant biomass and feeding the fermentation processes to produce 2G bioethanol. Wood-degrading fungi are an attractive source of these type of enzymes. Furthermore, as seen before, in addition to the glycosyl hydrolysis, these enzymes can support transglycosylation activity that can be of interest as an alternative to chemical glycosylation.

The glycosylation of bioactive compounds can increase the solubility[32] in aqueous media, improve the stability and biosafety[33], and make the new compound safer[34].

The BGLs employed in this chapter are secreted by the fungus *Talaromyces amestolkiae*. BGL-1 belongs to the GH1 family, while BGL-2 belongs to the GH3. In the following section, the transglucosylation properties of BGL-1 and BGL-2 of *Talaromyces amestolkiae versus* a panel of potential acceptors and the characterization of the products are presented.

6.3.2. Results and discussion

6.3.2.1. BGL-2

Hydroxytyrosol and vanillyl alcohol are phenolic compounds that present antioxidant properties. The poor bioavailability they present can be improved by glycosylation. In fact, the glucosylated derivatives of hydroxytyrosol and vanillyl alcohol have been reported to be more effective and safer when used against breast cancer cells[35].

The transglucosylation activity of BGL-2 (native) was tested using cellobiose (glucose β(1-4) glucose) as donor and hydroxytyrosol and vanillyl alcohol as acceptors.

As in the case of the xylose derivatives, the ^{1}H-^{13}C HSQC and HMBC experiments were employed to confirm the formation of the glucose derivatives and the position where the acceptors were glucosylated.

6.3.2.1.1. Hydroxytyrosol

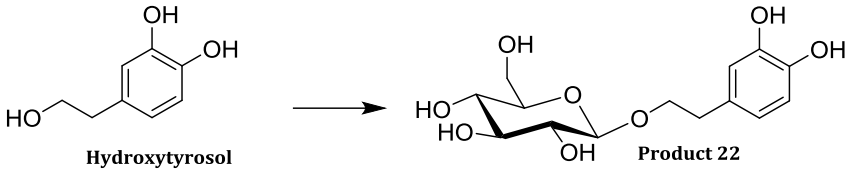


Figure 21. Product obtained by transglucosylation using hydroxytyrosol as acceptor.

The correlation observed in the HMBC (figure 22) between the anomeric carbon (C1) and the carbon C8' of the hydroxytyrosyl residue confirmed the formation of the derivative through the primary hydroxyl group of the aliphatic chain but not to the phenolic OH.

Moreover, the anomeric proton presented a coupling constant value of 8 Hz, pointing out the formation of the derivative through a β-linkage, showing the regioselectivity of the enzyme.

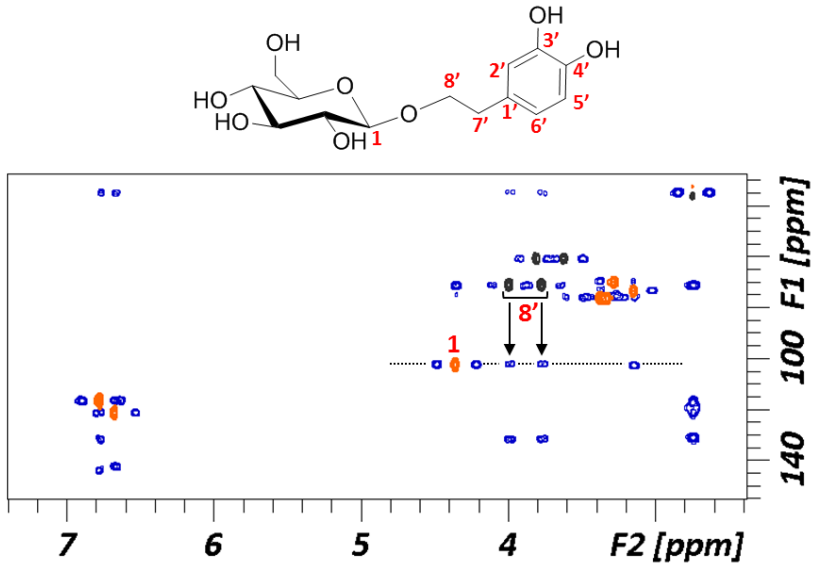


Figure 22. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the hydroxytyrosol derivative.

6.3.2.1.2. Vanillyl alcohol

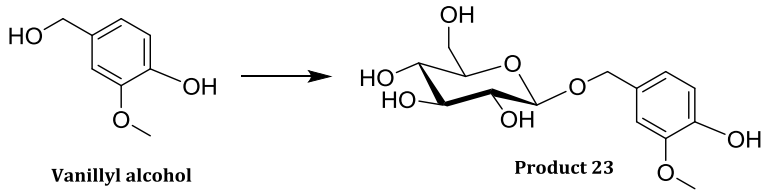


Figure 23. Product obtained by transglucosylation using vanillyl alcohol as acceptor.

As in the case of the hydroxytyrosol, the linkage through the aliphatic chain was demonstrated by the correlation observed in the HMBC (figure 24) between the anomeric carbon (C1) and the carbon C7' of the vanillyl residue.

Moreover, the coupling constant value of the anomeric proton was 7.7 Hz, pointing out the formation of the derivative through a β-linkage, and showing again the regioselectivity of the enzyme.

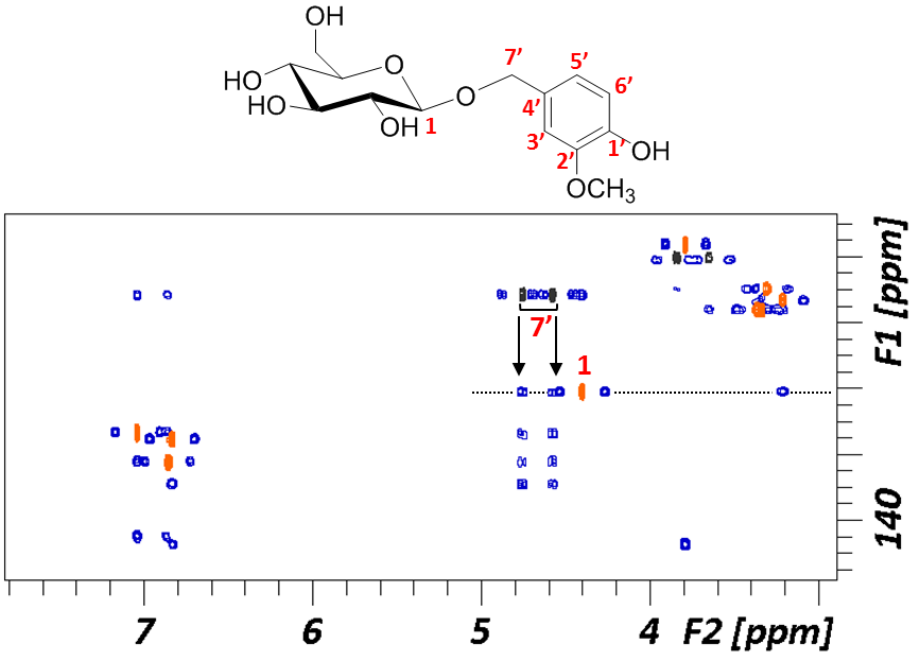


Figure 24. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the vanillyl alcohol derivative.

Interestingly, although the BGL-2 glucosidase and the BxTW1 xylosidase are both classified in the GH3 family, they have different selectivity towards alcohol groups. BGL-2 shows preference for aliphatic alcohols while the xylosidase prefers aromatic alcohols. Further studies are needed to understand those different behaviors in mechanistic and structural terms.

6.3.2.2. <u>BGL-1</u>

As previously mentioned, although GHs can be successfully used to synthesize glycoconjugates, the yields are often poor. The reaction proceeds under kinetic control and the newly-formed products are easily hydrolyzed by the same GH when the initial glycoside donor is exhausted, and thus the process is not economically viable at large-scales. In the way for developing improved biocatalyst for glycosylation reaction, BGL-1 from GH-1 was chosen to explore the "glycosynthase" strategy, mutating the nucleophilic carboxylate residue at the active site. In the case of BGL-1, it corresponds to glutamic 521. After testing several mutants, the variant E521G was selected. As the glycosynthase form is devoid of hydrolytic activity, an

activated donor is needed (α-fluoro-glucose in this case), mimicking the glycosyl-enzyme intermediate in α configuration (figure 25).

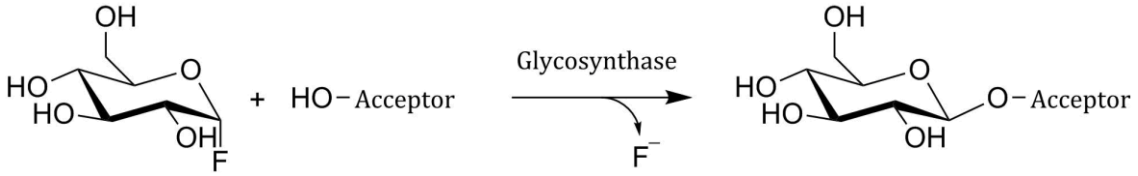


Figure 25. Transglycosidation reaction catalyzed by the glycosynthase type mutated enzyme.

In this case, β-glycosides that should be the canonical substrates of the hydrolytic activity of the native enzyme, are not hydrolyzed by the mutated enzyme. However, they can be used as potential saccharide acceptors to study the regioselectivity of transglycosylations. Two different type of acceptors were employed, *p*NP-glycosides and EGCG.

6.3.2.2.1. *p*NP-glycosides

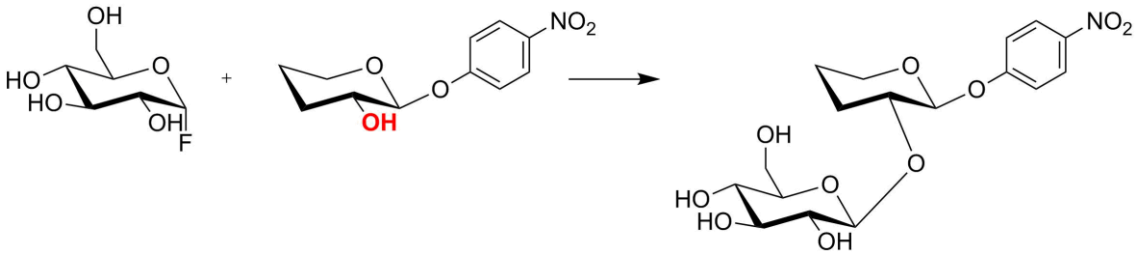


Figure 26. Reaction scheme of αF-glucose and the *p*NPX derivative acceptor.

*p*NP-glycosides of glucose, galactose and xylose were tested as acceptors.

Surprisingly, a β(1-2) linkage was observed in the three cases (figure 27). The HMBC showed a correlation between the anomeric position of the non-reducing carbohydrate (glucose) and the anchorage position at the carbohydrate linked to the *p*NP. Moreover, the anomeric proton of the transferred glucose presented a coupling constant value of *ca.* 7 Hz in the three cases, pointing out the formation of the disaccharide through a β-linkage.

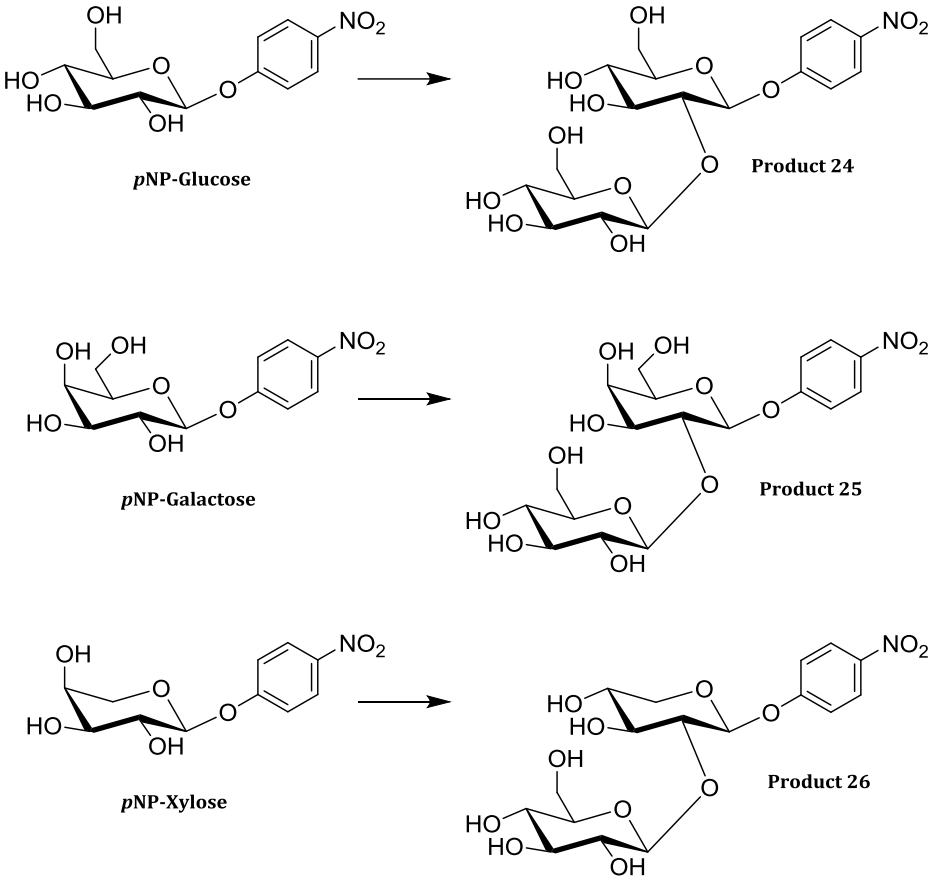


Figure 27. Products obtained by transglucosylation using *p*NP-glycosydes as acceptors.

As an example, the characterization of the sophorose-*p*NP (glucose-glucose-*p*NP, product **24**) is presented in figure 28. The two key connectivity correlation peaks, one between the H1 of the middle glucose residue with the C1'' of the nitrophenol ring and the other between the H1' of the non-reducing end glucose and the C2 of the middle glucose are highlighted, demonstrating the (1-2) regioselectivity of the transglycosylation. The glucose-galactose-*p*NP and glucose-xylose-*p*NP derivatives (products **25** and **26**) are described in the appendix 3.

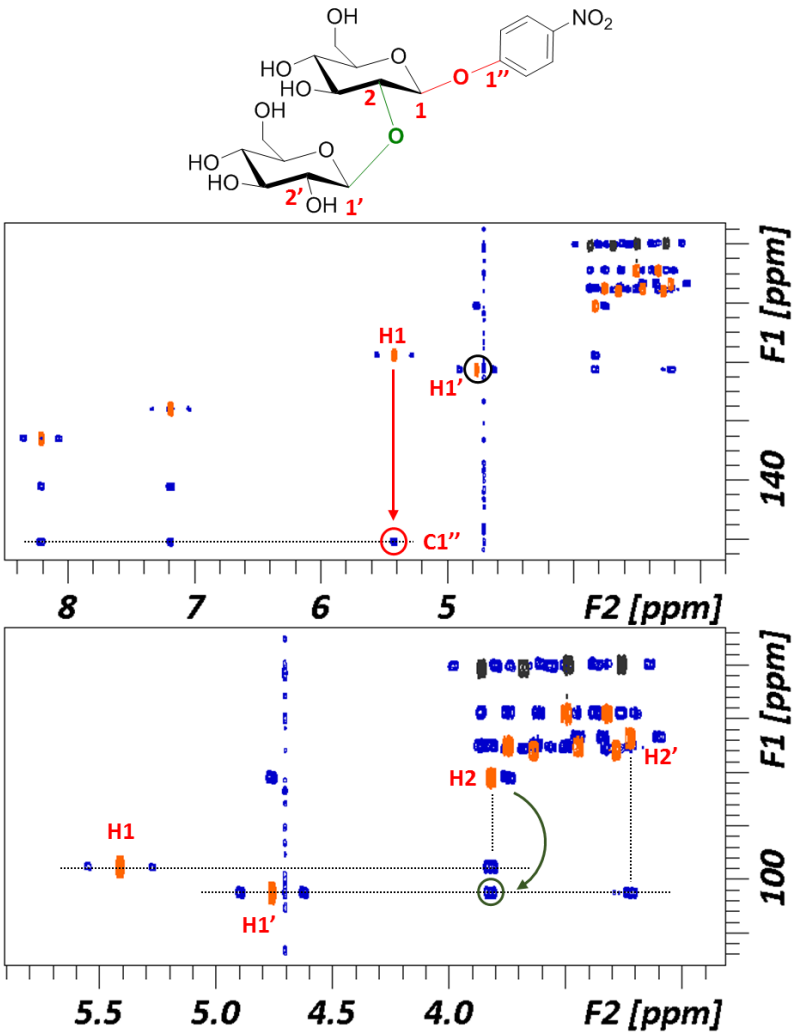


Figure 28. Top: Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the sophorose-*p*NP. **Bottom:** Expansion of the carbohydrate region. The arrows represent the key cross peaks for the characterization of the molecule.

6.3.2.2.2. Epigallocatechin gallate (EGCG)

After the reaction of α-fluoro-glucose and EGCG in the presence of E521G BGL-1 mutant, two compounds were obtained and submitted to HPLC analysis, which suggested the obtention of an EGCG-glucose derivative and another product with a second incorporation of glucose into the previously glucosylated EGCG.

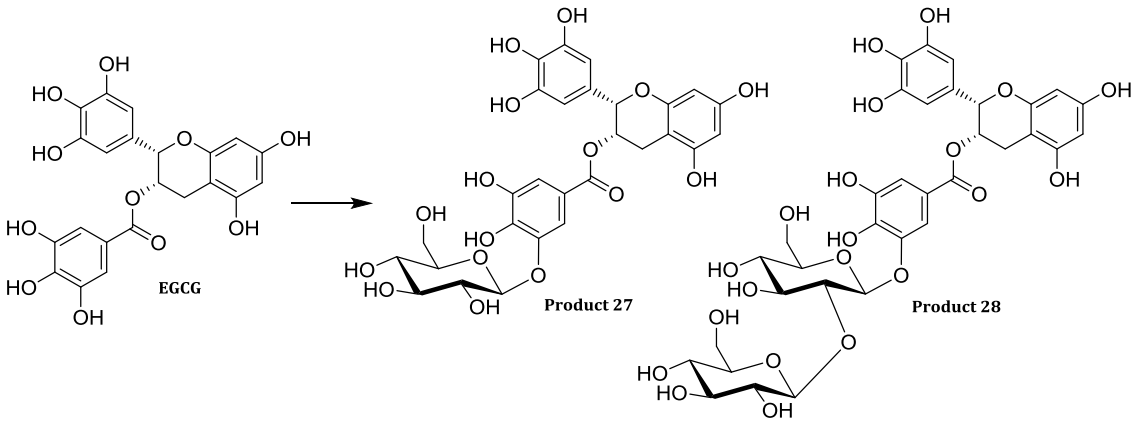


Figure 29. Products obtained by transglucosylation using EGCG as acceptor.

The same protocol described above was used to characterize both derivatives. The epigallocatechin gallate presents multiple -OH susceptible of glucosylation, reinforcing the needed of NMR as a tool to decipher the position of the second incorporated glucose.

The NMR analysis of the glucose derivative revealed the formation through a β-linkage (7.8 Hz coupling constant for the anomeric proton) to the gallate ring, determined by the correlation between the anomeric position of glucose (1) and the carbons 3'''/5''' of the gallate unit. Interestingly, the glycosylation differed from that observed for the xylose derivative, which took place through the carbon 4''' and also from the already mentioned α-glucosylated EGCG described by Plou et al.[29]

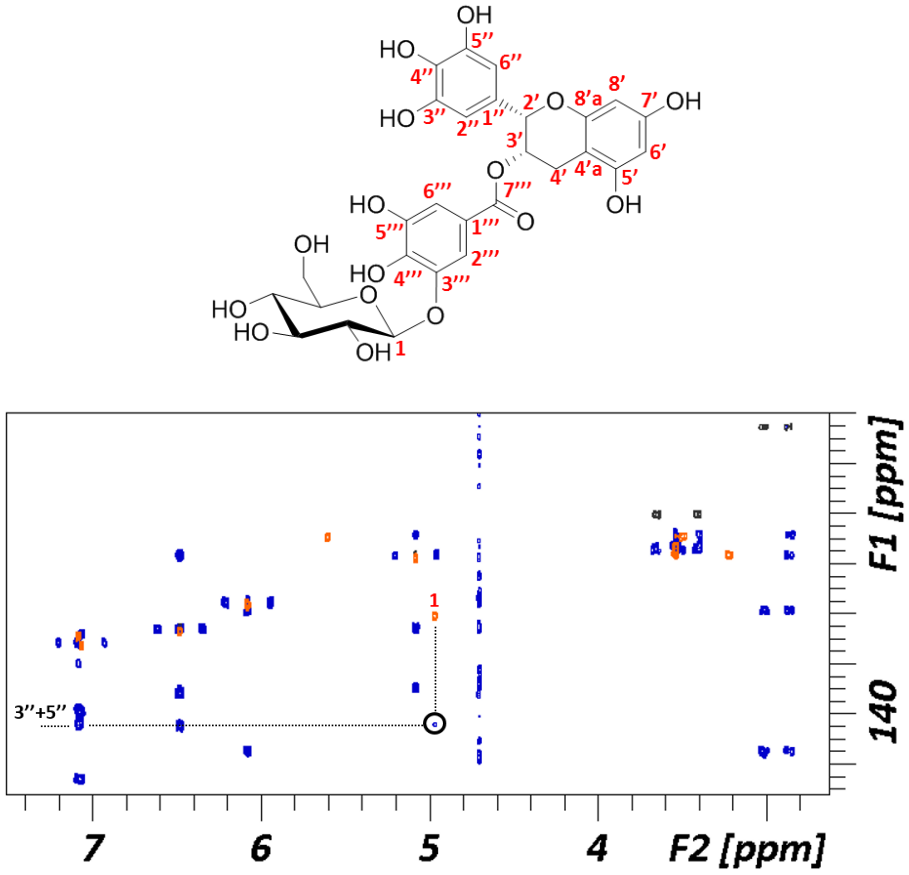


Figure 30. Superimposition of ^{1}H-^{13}C HSQC (black/orange) and HMBC (blue) spectra of the EGCG glucose derivative.

In the case of the product with a second incorporation of glucose, the NMR data pointed out the formation of a sophorose disaccharide moiety. Therefore, the second glucose ring was incorporated to the glucose residue of the derivative, instead of the glucosylation of a second hydroxyl group of the EGCG. Moreover, the coupling constant value of the anomeric proton was 7 Hz, pointing out again a β-linkage formation.

The transglucosylation by a β(1-2) linkage led to question if the mutation of the enzyme was the responsible of the regioselectivity observed or if the native enzyme already presented the β(1-2) regioselectivity. However, the native enzyme showed low efficiency in the transglycosylation path compared to the hydrolytic reaction, making difficult the study of the products obtained by transglucosylation catalyzed by the native enzyme.

To avoid the hydrolysis of the tranglycosylation product before being observable by NMR, a reaction was set using *p*NP-Glucose as donor and ^{13}C labelled glucose as

acceptor, with the aim of reducing the acquisition time of the ^{1}H-^{13}C HSQC NMR spectra to directly register them during the enzymatic reaction course. The ^{1}H-^{13}C HSQC demonstrated the transglucosylation by β(1-2) linkage, by observing the shifting of the C_2-H cross-peak of the labelled glucose to the position corresponding to sophorose.

The comparison of the ^{1}H-^{13}C HSQC spectra (figure 31) of commercial cellobiose, sophorose and laminaribiose (glucose β(1-4) glucose, glucose β(1-2) glucose and glucose β(1-3) glucose, respectively) with the product obtained in the reaction mixture, unequivocally demonstrated the unique formation of sophorose. Therefore, the β(1-2) regioselectivity also for the native BGL-1 was confirmed.

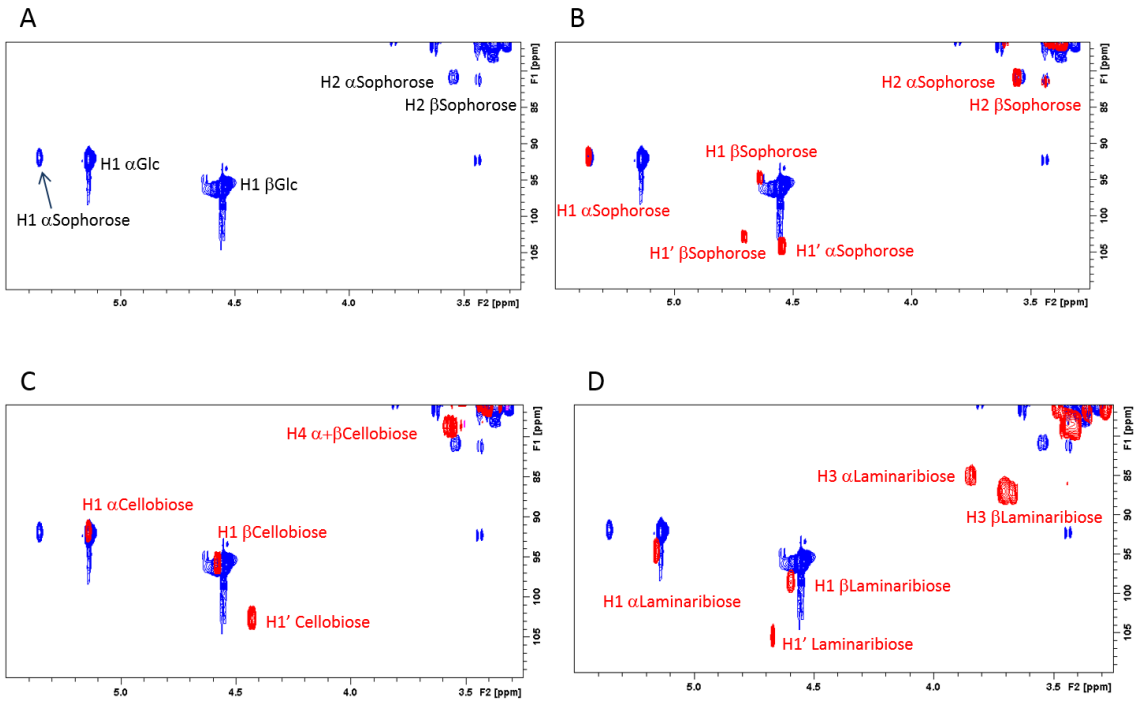


Figure 31. ^{1}H-^{13}C HSQC NMR spectra for detecting the transglycosylation regioselectivity of BGL-1. A) ^{1}H-^{13}C HSQC NMR spectra of the reaction mixture using *p*NPG as donor and ^{13}C labelled glucose as acceptor. The high intensity of the signals belonging to the excess of the ^{13}C-Glucose are due to the isotopic labelling. B) Superimposition of the ^{1}H-^{13}C HSQC NMR spectra of the reaction (in blue) and sophorose (in red). H1' of α and βSophorose are not observed in the reaction spectrum since the non-reducing residue comes from the donor (not ^{13}C labelled). Only signals coming from the ^{13}C-Glucose (acceptor) can be observed (short acquisition NMR time). C) Superimposition of the ^{1}H-^{13}C HSQC NMR spectra of the reaction (in blue) and cellobiose (in red). D) Superimposition of the ^{1}H-^{13}C HSQC NMR spectra of the reaction (in blue) laminaribiose (in red).

6.3.3. Experimental

The native and mutated enzymes, BGL-1 and BGL-2, the α-fluoro-glucose and the purified transglycosylation products were provided by the group of María Jesús Martinez (Biological Research Center in Madrid, CIB-CSIC). The glycosyl donors and acceptors were from commercial sources.

The structure and numeration of the derivatives can be found in the appendix 4.

NMR experiments of the purified compounds were acquired at 298 K using a Bruker AVANCE 500 MHz spectrometer and a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenic probe (magnetic field indicated in the characterization section). Samples were prepared in 500 µL of deuterated water (D_2O). 1D ^1H NMR spectra, ^1H-^{13}C HSQC, HMBC, ^1H-^1H TOCSY and selective TOCSY experiments were carried out in order to fully assign the compounds. Corresponding pulse sequences included in the TOPSPIN Bruker software were employed.

	^1H (δ ppm)	^{13}C (δ ppm)
	Product 22	
H_1	4.36	102.2
H_2	3.15	73.25
H_3	3.35	76.38
H_4	3.27	69.65
H_5	3.35	76.38
H_6	3.62	60.60
	3.81	
$H_{1'}$	---	131.79
$H_{2'}$	6.77	116.38
$H_{3'}$	---	143.81
$H_{4'}$	---	142.18
$H_{5'}$	6.77	116.38
$H_{6'}$	6.67	121.18
$H_{7'}$	2.74	35.74
$H_{8'}$	3.77	70.77
	3.99	

	Product 23	
H_1	4.39	100.94
H_2	3.20	73.02
H_3	3.35	75.94
H_4	3.30	69.61
H_5	3.35	75.94
H_6	3.65	60.31
	3.82	
$H_{1'}$	---	144.86
$H_{2'}$	---	147.46
$H_{3'}$	7.03	113.29

	^1H (δ ppm)	^{13}C (δ ppm)
	Product 26	
H_1	5.40	98.39
H_2	3.82	80.95
H_3	3.72	74.44
H_4	3.71	68.71
H_5	3.49	64.94
	3.98	
$H_{1'}$	4.74	102.91
$H_{2'}$	3.22	73.65
$H_{3'}$	3.44	75.36
$H_{4'}$	3.32	69.15
$H_{5'}$	3.28	75.89
$H_{6'}$	3.25	60.01
	3.48	
$H_{1''}$	---	161.22
$H_{2''}$	8.20	126.06
$H_{3''}$	7.17	116.13
$H_{4''}$	---	142.56
$H_{5''}$	7.17	116.13
$H_{6''}$	8.20	126.06

	Product 27	
H_1	4.97	100.33
H_2	3.53	72.60
H_3	3.54	75.32
H_4	3.49	68.65
H_5	3.22	75.97
H_6	3.41	59.73
	3.66	

$H_{4'}$	---	128.97
$H_{5'}$	6.85	121.36
$H_{6'}$	6.83	115.37
$H_{7'}$	4.56	71.39
	4.76	
$H_{8'}$	3.78	55.98

	Product 24	
H_1	5.41	97.9
H_2	3.82	81.29
H_3	3.74	75.28
H_4	3.49	68.98
H_5	3.63	76.13
H_6	3.68	60.71
	3.86	
$H_{1'}$	4.77	102.83
$H_{2'}$	3.22	73.88
$H_{3'}$	3.44	75.46
$H_{4'}$	3.32	69.23
$H_{5'}$	3.27	76.03
$H_{6'}$	3.26	60.20
	3.48	
$H_{1''}$	---	161.19
$H_{2''}$	8.20	126.15
$H_{3''}$	7.18	116.18
$H_{4''}$	---	142.52
$H_{5''}$	7.18	116.18
$H_{6''}$	8.20	126.15

	Product 25	
H_1	5.35	98.28
H_2	3.98	79.60
H_3	3.90	72.22
H_4	3.97	68.02
H_5	3.87	75.32
H_6	3.70	60.73
	3.70	
$H_{1'}$	4.74	103
$H_{2'}$	3.20	73.68
$H_{3'}$	3.45	75.42
$H_{4'}$	3.30	69
$H_{5'}$	3.25	75.94
$H_{6'}$	3.18	59.83
	3.46	
$H_{1''}$	---	161.43
$H_{2''}$	8.20	126.05
$H_{3''}$	7.19	116.18
$H_{4''}$	---	142.52
$H_{5''}$	7.19	116.18
$H_{6''}$	8.20	126.05

$H_{2'}$	5.08	77.06
$H_{3'}$	5.60	68.89
$H_{4'}$	2.86	24.91
	3.01	
$H_{4a'}$	---	99.04
$H_{5'}$	---	155.18
$H_{6'}$	6.08	95.72
$H_{7'}$	---	155.17
$H_{8'}$	6.08	95.72
$H_{8a'}$	---	155.18
$H_{1''}$	---	106.33
$H_{2''}$	6.48	106.45
$H_{3''}$	---	145.42
$H_{4''}$	---	132.10
$H_{5''}$	---	145.44
$H_{6''}$	6.48	106.45
$H_{1'''}$	---	120.59
$H_{2'''}$	7.09	108.61
$H_{3'''}$	---	144.27
$H_{4'''}$	---	139.73
$H_{5'''}$	---	144.30
$H_{6'''}$	7.06	112.27
$H_{7'''}$	---	166.50

	Product 28	
H_1	5.09	99.22
H_2	3.78	80.98
H_3	3.69	75.39
H_4	3.53	68.29
H_5	3.21	75.63
H_6	3.41	59.83
	3.66	
$H_{1'}$	4.77	102.77
$H_{2'}$	3.25	73.85
$H_{3'}$	3.43	75.70
$H_{4'}$	3.36	69.36
$H_{5'}$	3.29	76.09
$H_{6'}$	3.25	60.33
	3.53	
$H_{2''}$	5.03	77.05
$H_{3''}$	5.57	68.71
$H_{4''}$	2.84	24.86
	2.97	
$H_{4a''}$	---	99
$H_{5''}$	---	155.34
$H_{6''}$	6.06	96.01
$H_{7''}$	---	155.34
$H_{8''}$	6.06	96.01
$H_{8a''}$	---	155.21
$H_{1'''}$	---	106.39
$H_{2'''}$	6.47	106.40
$H_{3'''}$	---	145.51
$H_{4'''}$	---	132.62
$H_{5'''}$	---	145.65
$H_{6'''}$	6.47	106.40
$H_{1''''}$	---	120.58

$H_{2''''}$	7.06	108.61
$H_{3''''}$	---	144.59
$H_{4''''}$	---	139.90
$H_{5''''}$	---	144.59
$H_{6''''}$	7.05	112.18
$H_{7''''}$	---	166.43

Table 4. ^{1}H-^{13}C NMR chemical shift (δ, ppm) of the transglycosylation products **22** to **28** (see appendix 4 figure 38 for molecule structures and atom numbering).

Product 22: ^{1}H NMR (600 MHz, D_2O) δ 6.77 (d, *J* = 7.4 Hz, 2H), 6.67 (d, *J* = 8.2 Hz, 1H), 4.36 (d, *J* = 8.0 Hz, 1H), 4.00 (dt, *J* = 9.9, 7.0 Hz, 1H), 3.86 – 3.74 (m, 2H), 3.62 (dd, *J* = 12.3, 5.8 Hz, 1H), 3.43 – 3.24 (m, 3H), 3.15 (t, *J* = 8.6 Hz, 1H), 2.75 (t, *J* = 6.9 Hz, 1H).

Product 23: ^{1}H NMR (600 MHz, D_2O) δ 7.03 (d, *J* = 2.1 Hz, 1H), 6.91 – 6.77 (m, 2H), 4.75 (d, *J* = 11.5 Hz, 1H), 4.57 (d, *J* = 11.4 Hz, 1H), 4.40 (d, *J* = 8.0 Hz, 1H), 3.83 (dd, *J* = 12.5, 2.1 Hz, 1H), 3.78 (s, 3H), 3.64 (dd, *J* = 12.3, 5.8 Hz, 1H), 3.45 – 3.25 (m, 3H), 3.21 (t, *J* = 8.6 Hz, 1H).

Product 24: ^{1}H NMR (500 MHz, D_2O) δ 8.22 (d, *J* = 8.8 Hz, 2H), 7.19 (d, *J* = 8.8 Hz, 2H), 5.42 (d, *J* = 7.5 Hz, 1H), 4.77 (d, *J* = 8.0 Hz, 1H), 3.94 – 3.79 (m, 2H), 3.75 (t, *J* = 9.1 Hz, 1H), 3.67 (ddd, *J* = 20.8, 9.5, 5.1 Hz, 2H), 3.56 – 3.41 (m, 3H), 3.38 – 3.17 (m, 4H).

Product 25: ^{1}H NMR (500 MHz, D_2O) δ 8.22 (d, *J* = 8.8 Hz, 2H), 7.20 (d, *J* = 8.8 Hz, 2H), 5.37 (d, *J* = 7.4 Hz, 1H), 4.10 – 3.80 (m, 4H), 3.71 (d, *J* = 6.1 Hz, 2H), 3.56 – 3.38 (m, 2H), 3.40 – 3.11 (m, 4H).

Product 26: ^{1}H NMR (500 MHz, D_2O) δ 8.21 (d, *J* = 8.8 Hz, 2H), 7.19 (d, *J* = 8.8 Hz, 2H), 5.42 (d, *J* = 6.9 Hz, 1H), 3.99 (dd, *J* = 11.9, 3.9 Hz, 1H), 3.83 (t, *J* = 7.5 Hz, 1H), 3.72 (dd, *J* = 6.6, 4.0 Hz, 2H), 3.48 (ddd, *J* = 26.6, 14.1, 7.2 Hz, 3H), 3.39 – 3.16 (m, 4H).

Product 27: ^{1}H NMR (600 MHz, D_2O) δ 7.13 – 7.02 (m, 2H), 6.48 (s, 2H), 6.08 (s, 2H), 5.60 (s, 1H), 5.08 (s, 1H), 5.01 – 4.92 (m, 1H), 3.65 (dd, *J* = 12.8, 3.5 Hz, 1H), 3.58 – 3.16 (m, 5H), 3.01 (dd, *J* = 17.3, 4.6 Hz, 1H), 2.86 (d, *J* = 18.5 Hz, 1H).

Product 28: ^{1}H NMR (600 MHz, D_2O) δ 7.06 (s, 2H), 6.47 (s, 2H), 6.06 (q, *J* = 2.4, 2.0 Hz, 2H), 5.60 – 5.54 (m, 1H), 5.09 (d, *J* = 7.6 Hz, 1H), 5.03 (s, 1H), 4.78 (d, *J* = 8.0 Hz, 1H), 3.78 (t, *J* = 8.4 Hz, 1H), 3.73 – 3.63 (m, 2H), 3.57 – 3.50 (m, 2H), 3.47 – 3.39 (m, 2H), 3.36 (t, *J* = 9.4 Hz, 1H), 3.33 – 3.18 (m, 4H), 2.97 (dd, *J* = 17.6, 4.5 Hz, 1H), 2.84 (d, *J* = 18.5 Hz, 1H).

6.4. CONCLUSIONS

The study of the mechanism of action of three Glycosyl Hydrolase enzymes, xylosidase BxTW1 classified in the GH3 family, BGL-1 in GH1 and BGL-2 in GH3, obtained from the wood degrading fungi *Talaromyces amestolkiae*, and their transglycosylation activities have been performed.

The hydrolytic mechanism of native BxTW1 has been elucidated by NMR, showing a retaining mechanism. Moreover, the oligosaccharide hydrolysis selectivity has been determined following the signals of the ^{1}H NMR spectra recorded at different times in the course of the enzymatic reaction, observing the preference of xylotriose over xylobiose consumption. Furthermore, DOSY-NMR has allowed to determine the presence of transglycosylation products in the reaction mixture and the transxyloxylation activity has been confirmed when using recombinant rBxTW1.

The glycosidic bond synthesis activity has been exploited by using a thioglycosynthase type mutant, rBxTW1-E495A, demonstrating its ability to catalyze the synthesis of an extensive panel of xylosylated products with different glycosidic bonds (glycosyl phenols, glycosyl esters, glycosyl phosphate esters, tio- and seleno-glycosydes and N-glycosides). They have been fully characterized by NMR employing the ^{1}H-^{13}C HSQC and HMBC experiments.

The β(1-2) regioselectivity and transglycosylation products obtained with BGL-1 and its thioglycosynthase type mutant E521G have been demonstrated and characterized by NMR.

The transglycosylation products obtained with BGL-2 have been characterized by NMR, showing preference for aliphatic over aromatic alcohols.

6.5. REFERENCES

[1] G. Davies, B. Henrissat, *Structure* **1995**, *3*, 853–859.
[2] C. S. Rye, S. G. Withers, *Curr. Opin. Chem. Biol.* **2000**, *4*, 573–580.
[3] D. E. Koshland, *Biol. Rev.* **1953**, *28*, 416–436.
[4] M. L. Sinnott, *Chem. Rev.* **1990**, *90*, 1171–1202.
[5] A. White, D. R. Rose, *Curr. Opin. Struct. Biol.* **1997**, *7*, 645–651.
[6] T. D. Heightman, A. T. Vasella, *Angew. Chemie Int. Ed.* **1999**, *38*, 750–770.

[7] B. P. Rempel, S. G. Withers, *Glycobiology* **2008**, *18*, 570–586.
[8] P. Bojarová, V. Křen, *Trends Biotechnol.* **2009**, *27*, 199–209.
[9] J. Edelman, in *Adv. Enzymol. Relat. Subj. Biochem.*, John Wiley & Sons, Ltd, **1956**, pp. 189–232.
[10] D. H. Crout, G. Vic, *Curr. Opin. Chem. Biol.* **1998**, *2*, 98–111.
[11] B. Bissaro, P. Monsan, R. Fauré, M. J. O'Donohue, *Biochem. J.* **2015**, *467*, 17–35.
[12] P. M. Danby, S. G. Withers, *ACS Chem. Biol.* **2016**, *11*, 1784–1794.
[13] R. Lopez, A. Fernandez-Mayoralas, *J. Org. Chem.* **1994**, *59*, 737–745.
[14] J. J. Aragón, F. J. Cañada, A. Fernández-Mayoralas, R. López, M. Martín-Lomas, D. Villanueva, *Carbohydr. Res.* **1996**, *290*, 209–216.
[15] Y.-S. Kim, C.-S. Park, D.-K. Oh, *Enzyme Microb. Technol.* **2006**, *39*, 903–908.
[16] M. Jahn, S. G. Withers, *Biocatal. Biotransformation* **2003**, *21*, 159–166.
[17] C. Malet, A. Planas, *FEBS Lett.* **1998**, *440*, 208–212.
[18] L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
[19] M. Jahn, J. Marles, R. A. J. Warren, S. G. Withers, *Angew. Chemie Int. Ed.* **2003**, *42*, 352–354.
[20] H. V. Scheller, P. Ulvskov, *Annu. Rev. Plant Biol.* **2010**, *61*, 263–289.
[21] R. Deutschmann, R. F. H. Dekker, *Biotechnol. Adv.* **2012**, *30*, 1627–1640.
[22] M. Nieto-Domínguez, L. I. de Eugenio, J. Barriuso, A. Prieto, B. F. de Toro, Á. Canales-Mayordomo, M. J. Martínez, *Appl. Environ. Microbiol.* **2015**, *81*, 6380–6392.
[23] M. Nieto-Domínguez, A. Prieto, B. Fernández de Toro, F. J. Cañada, J. Barriuso, Z. Armstrong, S. G. Withers, L. I. de Eugenio, M. J. Martínez, *Microb. Cell Fact.* **2016**, *15*, 171.
[24] M. Nieto-Domínguez, J. A. Martínez-Fernández, B. F. de Toro, J. A. Méndez-Líter, F. J. Cañada, A. Prieto, L. I. de Eugenio, M. J. Martínez, *Microb. Cell Fact.* **2019**, *18*, 174.
[25] P. Groves, M. O. Rasmussen, M. D. Molero, E. Samain, F. J. Cañada, H. Driguez, J. Jiménez-Barbero, *Glycobiology* **2003**, *14*, 451–456.
[26] K. Mani, M. Belting, U. Ellervik, N. Falk, G. Svensson, S. Sandgren, F. Cheng, L.-A. Fransson, *Glycobiology* **2004**, *14*, 387–397.
[27] M. Jacobsson, U. Ellervik, M. Belting, K. Mani, *J. Med. Chem.* **2006**, *49*, 1932–1938.
[28] S. Quideau, D. Deffieux, C. Douat-Casassus, L. Pouységu, *Angew. Chemie Int. Ed.* **2011**, *50*, 586–621.
[29] J. L. Gonzalez-Alfonso, L. Leemans, A. Poveda, J. Jimenez-Barbero, A. O. Ballesteros, F. J. Plou, *J. Agric. Food Chem.* **2018**, *66*, 7402–7408.
[30] T. Bravman, V. Belakhov, D. Solomon, G. Shoham, B. Henrissat, T. Baasov, Y. Shoham, *J. Biol. Chem.* **2003**, *278*, 26742–9.
[31] K. Kohata, E. Ohtaki, C. Suzuki, H. Meguro, *Nippon Kagaku Kaishi* **1982**, *1982*, 1609–1616.
[32] P. Torres, A. Poveda, J. Jimenez-Barbero, J. L. Parra, F. Comelles, A. O. Ballesteros, F. J. Plou, *Adv. Synth. Catal.* **2011**, *353*, 1077–1086.
[33] H.-J. Woo, H.-K. Kang, T. T. H. Nguyen, G.-E. Kim, Y.-M. Kim, J.-S. Park, D. Kim, J. Cha, Y.-H. Moon, S.-H. Nam, et al., *Enzyme Microb. Technol.* **2012**, *51*, 311–318.

[34] V. Kren, L. Martinkova, *Curr. Med. Chem.* **2001**, *8*, 1303–1328.
[35] J. A. Méndez-Líter, I. Tundidor, M. Nieto-Domínguez, B. F. de Toro, A. González Santana, L. I. de Eugenio, A. Prieto, J. L. Asensio, F. J. Cañada, C. Sánchez, et al., *Microb. Cell Fact.* **2019**, *18*, 97.